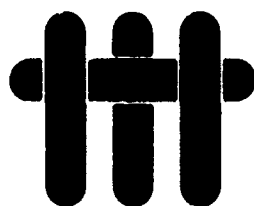# ANNUAL REPORT
University Research Initiative
→ Contract No.: N00014-86-K-0753
September 15, 1988-September 14, 1989

The Processing and Mechanical Properties of High Temperature/ High Performance Composites

by

A. G. Evans & R. Mehrabian
University of California,
Santa Barbara

Arizona State University
Cambridge University
Harvard University
Stanford University
Washington State University
University of Wisconsin-Madison

Book 1 of 6
Section 1: COATINGS and INTERFACES

89 12 21 035

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION	1b RESTRICTIVE MARKINGS
Unclassified	

2a. SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION/AVAILABILITY OF REPORT
	See Distribution List
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	

4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)

6a. NAME OF PERFORMING ORGANIZATION	6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
University of California Materials Department		Office of Naval Research

6c. ADDRESS (City, State, and ZIP Code)	7b. ADDRESS (City, State, and ZIP Code)
College of Engineering University of California Santa Barbara, CA 93106	800 North Quincy Street Arlington, VA 22217-5000

8a. NAME OF FUNDING/SPONSORING ORGANIZATION	8b. OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
Defense Advanced Research Projects Agency		#N00014-86-K-0753

8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS			
1400 Wilson Blvd. Arlington, VA 22209	PROGRAM ELEMENT NO	PROJECT NO.	TASK NO	WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
The Processing and Mechanical Properties of High Temperature/High Performance Composites

12. PERSONAL AUTHOR(S)
Evans, Anthony G. and Mehrabian, Robert

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Annual	FROM 880915 TO 890914	891015	2062

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Interfaces, Coatings, Toughening, Processing, Composites, Ceramics, Intermetallics

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The third annual report of the University Research Initiative project at UCSB on High-Temperature, High-Performance Composites consists of four sections compiled in a total of six books. The first sections is concerned with coatings and interfaces and is expressed in Book 1. The second section focuses on strength and fracture resistance and is found in Books 2 and 3. The third section describes research on flow and creep strength, this information is presented in Book 4. The fourth section is concerned with various aspects of processing in matrices and composites and is described in Books 5 and 6.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL

DD Form 1473, JUN 86   Previous editions are obsolete.

SUMMARY

of

TABLE OF CONTENTS

EXECUTIVE SUMMARY

BOOK 1

SECTION 1: COATINGS AND INTERFACES

BOOK 2

SECTION 2: STRENGTH AND FRACTURE RESISTANCE

Part 1 of 2

Ductile Reinforcement

BOOK 3

SECTION 2: STRENGTH AND FRACTURE RESISTANCE

Part 2 of 2

BOOK 4

SECTION 3: FLOW AND CREEP STRENGTH

BOOK 5

SECTION 4: PROCESSING: *Matrices and Composites*

Part 1 of 2

Intermetallics

BOOK 6

SECTION 4: PROCESSING: *Matrices and Composites*

Part 2 of 2

Ceramics

EXECUTIVE SUMMARY

The third annual report of the University Research Initiative project at UCSB on High-Temperature, High-Performance Composites consists of sections compiled in a total of six books. The first section in Book 1 is concerned with the properties and structure of bimaterial *interfaces* and the related problem of *coating* decohesion and cracking. The second section describes research on the *strengths and fracture resistance* of brittle matrix composites manufactured with fibers, whiskers and ductile phases. This information is presented in Books 2 and 3. The third section addresses the *flow and creep strength* of reinforced systems, with emphasis on effects of aspect ratio and the incidence of damage, and is offered in Book 4. The fourth section, Books 5 and 6, describes work on processing of intermetallic and ceramic matrices and composites, as well as numerical modelling of the melt-spinning process.

SECTION 1: COATINGS AND INTERFACES

The research on *interfaces* has utilized basic mechanics concepts for interface cracks for the continued development of test methods for measuring the fracture resistance properties of interfaces in the range of mode mixity of most interest. These methods, coupled with previously devised tests, have been used to obtain data for numerous bimaterial systems. A particular emphasis has been the acquisition of data for potential *fiber coating* systems. The results have revealed that various oxide coatings (such as Y_2O_3 and Al_2O_3) have low fracture energies within refractory alloy toughened intermetallics and thereby encourage extensive debonding, as elaborated below. For oxide based composites, not surprisingly, most oxide coatings provide a fracture energy unacceptably large for debonding

purposes. However, ZrO_2 coatings on Al_2O_3 are an exception; they exhibit a low fracture energy and debond extensively. These results provide guidance for a related program on fiber coatings in composites.

The overall mechanical behavior of *external coatings* have been explored, with emphasis on coating decohesion and on the role of ductile inclusions in the cracking of the coating. The ductile inclusions are found reduce the cracking property of the coating, as evident for cracking patterns. The mechanics of cracks at coating interfaces subject to combinations of mode I, II and III has also been formulated (Jensen et al.), leading to test procedures for readily examining the effects of modes II and III on the interface decohesion energy. A simple formulation for representing effects of the shear modes on the interface decohesion energy has also been proposed and shown to provide a good description of available data.

SECTION 2: STRENGTH AND FRACTURE RESISTANCE

FIBER REINFORCEMENT

The axial tensile properties of a range of fiber reinforced ceramics have been rigorously evaluated and shown to be consistent with models previously developed in the program. In particular, the matrix cracking stresses and the ultimate strength have been predicted through the models based on independent measurements of the elastic properties, the interface sliding stress, the interface debond energy, the residual strain and the *in situ* strengths of the fibers. The basic models applicable to tensile properties, as well as the methods for measuring the important microstructural properties, are thus concluded to capture the essential features of composite behavior.

Mode I resistance curve measurements have been made on composites heat treated to produce fiber interfaces having sufficient sliding resistance that

delamination cracking is suppressed (Sbaizero). The materials and test specimens lead to large scale bridging and the measured data require correction for this, as elaborated below. The crack growth behavior has also been numerically simulated (Hom and McMeeking) based on bridging and sliding traction laws derived in previous research within the program. Good agreement between the simulation and experimental data is demonstrated for an interface sliding stress consistent with the measured fiber pull-out lengths. The results also reveal that frictional dissipation during pull-out provides the main contribution to the toughness ($K_c \approx 20\,MPa\sqrt{m}$ at steady-state) and that large scale bridging effects lead to substantial overestimates of actual toughness levels when conventional linear elastic fracture mechanics formulae are used.

Delamination cracking has been investigated in a laminated composite (Sbaizero et al.) and the data have been interpreted based on solutions for mixed mode cracking in anisotropic media (Suo). Delamination crack growth resistances have been deduced and shown to be governed primarily by the matrix fracture energy, with some contribution from distributed fibers that bridge the crack surfaces. The crack is also found to progress into a steady-state trajectory along a laminate interface; furthermore, the initial crack path rotates toward that interface in a sense governed by the sign of K_{II}.

To further understand these effects, a new method for fabricating ceramic composites has been invented (Folsom et al.). The method involves bonding together thin ceramic sheets and thin layers of fibers to form a multiple sandwich composite comprising of alternate layers of ceramic and fibers. The thin, strong ceramic sheets are formed by a method used to make ceramic substrates for electronic packaging. The fiber layers can be in the form of either aligned fibers or cloth, which are penetrated with an appropriate bonding agent, e.g., an epoxy resin, metal or ceramic powder. The composite is formed by sandwiching the ceramic

3

sheets with fiber layers and then bonding the ceramic sheets to the fiber layers with an applied pressure at the appropriate temperature. Current work emphasizes composites formed with epoxy resin, carbon fiber prepregs laminated between ceramic sheets comprised of either alumina, transformation toughened zirconia, or glass by hot-pressing. Because processing is simple and carried out at low temperatures, large numbers of composite modifications are being explored.

Fracture resistance caused through bridging by fiber and/or by ductile reinforcements has been addressed in the large scale bridging regime (Zok and Hom), by experiments conducted on metal reinforced ceramics (Velamakanni et al.), on fiber reinforced ceramics (Sbaizero et al.) and on a metal fiber reinforced polymer. The results establish a strong influence of large scale bridging on the apparent fracture resistance, consistent with numerical simulations of crack growth in such materials. An approximate analytical solution has also been developed that correlates well with the data and may be used for interpretation purposes and, furthermore, generates actual material resistances from the experimental results. Additionally, *this work clearly indicates that nominal fracture toughness results can substantially overestimate actual properties and must be used with caution.*

DUCTILE REINFORCEMENT

Following the results of the previous year which highlighted the importance of interface debonding and of the reinforcement ductility (Ashby et al.) on the toughness, a systematic study has been completed on the effects of those variables (Cao et al.; Dève et al.). For this purpose, experiments have been conducted on composite cylinders and on laminated systems consisting of TiAl and Nb/Ti alloys (thickness, 2R), with various thin oxide coatings and/or reaction product layers between the matrix and reinforcements. The results reveal that Y_2O_3 coatings inhibit reactions and also allow extensive debonding (debond length d ≈ 20R), consistent

with its relatively low fracture energy ($\Gamma_i \approx 25 \, Jm^{-2}$). For high ductility reinforcements, such as high purity Nb, debonding leads to a large work of rupture ($\chi \to 6-7$). In the absence of a coating, more limited debonding occurs ($d \approx 5-6R$) along a σ-phase reaction product layer, resulting in a smaller work of rupture ($\chi \approx 2.5$). Furthermore, for the latter, χ is found to be similar for composite cylinders, for laminates and for actual composites. High strength, low ductility reinforcements, such as Ti-33% Nb, indicate different characteristics. In this case, extensive debonding (induced by Y_2O_3 coatings) results in abrupt reinforcement rupture and small χ (1–1.5). However, when debonding is suppressed by averting the formation of a brittle reaction products layer, inhomogeneous deformation in the Ti/Nb apparently suppresses necking and allows a moderately large work of rupture ($\chi \approx 2-3$ in laminates and $\chi \approx 4-5$ in composites). The approach needed to achieve optimum toughening thus depends sensitively on the flow and fracture behavior of the reinforcement. An important effect of matrix crack offset on χ has also been found (Ashby) in the sense that χ increases as the offset angle increases. It remains to ascertain how this benefit can be encouraged in actual composites.

An essential, related aspect of this study concerns diffusion-couple annealing experiments which have been used to explore interactions between γ-TiAl matrices and ductile-phase reinforcements based on β-(Ti,Nb) alloys. The nature and rate of evolution of the interfacial layers have been characterized and the interdiffusion coefficients of Nb and Al in the β-(Ti,Nb) phase have been calculated for the 900-100°C temperature range using a Matano-Boltzmann analysis. While the interdiffusion coefficients are dependent on alloy composition, the activation energies were found to be quite similar in all cases, about 250 ± 40 kJ/mole (Jewett et al.).

A process for producing ceramics with an interpenetrating network of a metallic alloy has been invented (Velakammani et al.). In this process, ceramic

powder is packed within a pyrolyzable preform, e.g., an organic fiber felt, by pressure filtration. The preform is then pyrolyzed at low temperature of produce a powder compact containing pore channels remnant of the preform. The ceramic powder is densified at higher temperatures without eliminating the pore channels which at infiltrated with a molten metal. A systematic study of the fracture toughness based on the above variables, which can be changed with this new processing method, is in progress. Initial studies have emphasized Al_2O_3 and transformation toughened ZrO_2 matrices containing 20 μm fibers of a Al-Mg alloy with an architecture remnant of a needled felt used to create the channels within the ceramic. Squeeze casting is used to intrude the molten metal into the pore channels. Preliminary fracture mechanics testing to determine crack growth resistance as a function of crack length has shown that the ductile bridging ligaments increase fracture toughness.

Studies of the deformation behavior of the reinforcements have provided additional insight. Deformation of TiNb involves rafted dislocation pile-ups in narrow, coarsely distributed slip bands that extend large distances away from the crack surfaces. Fracture occurs at the intersection of these bands. In the Nb system, twinning and slip occur in the γ matrix around the interface and near the crack plane. Furthermore, debonds along the σ phase reaction product layers appear to initiate at sites where the twins intersect this layer:

MATRIX TOUGHENING

An investigation of ceramics toughened with whiskers (Campbell et al.) has established both the prevalent contributions to toughness, as well as the realistic toughening potential. The two principal toughening contributions derive from the extra surface energy associated with debonding along the amorphous phase at the interface and the energy dissipated as acoustic waves when the whiskers fail in the

6

crack wake. These contributions can lead to toughness of order $K_c = 10 \, MPa \sqrt{m}$. Much larger toughness could be induced by encouraging frictional dissipation by sliding and pull-out along debonded interfaces. Direct measurements of these effects (Ashby) produce an opportunity to understand how this contribution can be understood and emphasized.

Fracture resistance effects have also been explored for process zone toughening mechanisms (Stump and Budiansky; Hom and McMeeking). These results show that the resistance curves exhibit a peak preceding steady-state and that the peak height is related to the gradient in volume fraction of transformed material in the process zone, $f(y)$, in the sense that uniform transformation in the zone leads to the maximum peak height. A comparison of simulated fracture resistance curves with experimental results (Hom and McMeeking) reveals good agreement when independent measurements of $f(y)$ and of the process zone size and shape are used to set the magnitudes of the parameters used in the simulation.

The overall toughness of a reinforced system may involve multiplicative effects between matrix toughening and reinforcement toughening. Matrix toughening behaviors include the transformation and whisker mechanisms noted above, as well as twin toughening. Multiplicative effects with reinforcement toughening occurs primarily with process zone mechanisms (twinning, transformation, etc.). Analysis of the coupled toughening (Stump and Budiansky, Hom and McMeeking) have demonstrated conditions that provide the extremes of multiplicative and additive behavior. Also, resistance curves applicable when multiple mechanisms operate have been simulated. These calculations provide the insight needed to select matrix microstructures consistent with that reinforcement scheme of choice.

SECTION 3: FLOW AND CREEP STRENGTH

A substantial activity has been initiated to examine strength and ductility during both plastic flow and creep in the presence of reinforcements. The variables of principal interest are the reinforcement aspect ratio and size, as well as their fracture/debond resistance. The initial studies have been on a model, solute-strengthened system, Al 4% Mg reinforced with SiC having different morphologies, fabricated by squeeze casting. The matrix is chosen to be ductile and fails by necking to a ridge and thereby, inhibits premature rupture.

Strengthening and ductility have been explored in the alloy containing equiaxed SiC particles (50% by volume) having a range of particle sizes between 3 and 160μm (Yang et al.). The results indicate that the flow strength diminishes somewhat with increase in particle size; this effect has been attributed to the enhanced damage (particle cracking) observed in materials containing the large particles. Conversely, there is a strong effect of particle size on ductility, with the greatest ductility occurring in materials containing the smallest particles. This trend is attributed to the relative incidence of particle cracking, as governed by weakest link statistics associated with flaws in the particles, introduced during comminution. The ultimate strengths of these materials is high (~ 600 MPa) and greater than that expected from calculations conducted for composites with a spatially uniform particle distribution. The discrepancy is being addressed by examination of the influence on non-uniform spatial arrangements and the associated high constraint between the more closely-spaced particles. Similarly large strength elevation has also been found for material reinforced with Al_2O_3 particulates (Hirth et al.). In this case, the strengthening has been attributed to dislocation cell formation governed by the particle spacing. These different effects remain to be brought together in a unified model. In addition, the toughness of the Al_2O_3 containing material has been

measured (Hirth et al.) to reveal that the toughness increases with increase in particle size in accordance with known concepts of ductile fracture.

SECTION 4: PROCESSING: Matrices and Composites

INTERMETALLICS

PHASE EQUILIBRIA - The assessment of the binary TiAl phase diagram has been completed through combined efforts on thermodynamic modelling, differential thermal analysis and annealing experiments (Mishurda et al.), as well as high-temperature X-ray diffractometry and controlled solidification experiments (McCullough et al.). In particular, the liquid-solid equilibrium in the vicinity of the equiatomic composition and the boundaries of the high temperature α-(Ti) field are now well-established. The revised diagram and corollary thermodynamic information has been instrumental not only in elucidating the microstructural evolution paths of binary Ti-Al intermetallic alloys, but also in permitting the systematic evaluation and computer calculation of related ternary systems, notably Ti-Al-Nb, Ti-Al-Ta and Ti-Al-B.

Substantial progress has been made on the Ti-Al-Nb ternary system. Extensive solidification experiments have been combined with the binary diagram information to define the approximate shape of the liquidus surface and the boundaries of the different primary phase fields (Jewett et al.). The α-(Ti) liquidus on the Ti-Al binary shrinks with the addition of Nb and is totally encroached by the β and γ-(TiAl) fields at ~ 18at%Nb. All other regions on the liquidus surface have temperatures that decrease from the Nb-Al binary toward the Ti-Al binary. Significant problems still exist in the vicinity of the Ti_2NbAl composition, where two ternary phases (T1 and T2) had been detected (Jewett et al., Boettinger et al.). Both

phases have B2 structures at 1200°C and decompose upon cooling to rather complex microstructures involving α_2, γ and a BB$_2$ phase.

Phase equilibria studies on the Ti-Al-Ta system have been initiated. Solidification work suggests that the shape of the liquidus surface in the vicinity of γ is quite similar to that reported for Ti-Al-Nb with the α -(Ti) field shrinking with Ta additions until it is encroached by the β -(Ti) and γ-(TiAl) fields (McCullough et al.), but the liquidus temperatures are generally higher than those for Nb additions. High temperature X-ray diffraction has been extensively used to study the phases present in the vicinity of the 50at%Al isoconcentrate (McCullough et al.). In addition, a preliminary isothermal section at 1100°C has been delineated based on the X-ray evidence combined with diffusion couples and isothermal annealing experiments developed under other DARPA sponsored programs.

SOLIDIFICATION - The fundamental understanding of the solidification of Ti-Al alloys developed in the previous funding period was systematically applied to elucidate the microstructure evolution of rapidly solidified powders produced by the industrial participants (McCullough et al., Jewett et al.). Microstructures of powders produced by containerless processing (PREP) were in general consistent with the trends established by the controlled solidification experiments, wherein β was found to be the kinetically preferred primary phase at high supercoolings, even in the composition range where α and/or γ are thermodynamically favored (45-55at%Al). In contrast, atomization processes involving graphite crucibles and/or nozzles (XSR) were found to yield powders with a large proportion of primary α-(Ti) phase, even at compositions where β is the equilibrium primary phase. This was ascribed to carbon contamination of the melt ($\sim$ 0.5at%C), which seems to shift the α/β liquidi boundary (McCullough et al.). Furthermore, the solidification behavior of dilute ($\geq$ 5at%Nb or Ta) ternary alloys in the $\alpha_2 - \gamma$ range was found to

proceed essentially as in the binary Ti-Al alloy of the same aluminum content (Jewett et al., McCullough et al.).

Studies on the effects of cooling rate after solidification using binary and dilute ($\geq$ 5at%Nb or Ta) ternary alloys (McCullough et al., Jewett et al.), revealed that diffusion-controlled lath transformation is sensitive to quenching. Indeed, γ precipitation may be completely suppressed by decreasing the Al content or adding ternary elements (Nb or Ta) that exhibit strong partitioning between the α_2 and γ phases. It was also found that the $\alpha \rightarrow \alpha_2$ ordering is not suppressible by quenching.

The solidification of γ-TiAl base alloys containing 10 to 28at%Ta (McCullough et al.) revealed that primary β-(Ti) forms with γ segregates in the interdendritic spaces. The leaner alloys appear to go through the double peritectic cascade previously reported for the binary system L + $\beta \rightarrow \alpha$ and L + $\alpha \rightarrow \gamma$, but the extent of α formation is reduced with increasing Ta content and most probably eliminated in the richer alloys. All these alloys exhibit strong partitioning of Ta toward the dendrite cores. This severe "coring" would require long homogenization treatments and its presence may lead to erroneous conclusions when mechanical and oxidation behavior studies are conducted on as-cast materials. Furthermore, the Ta-rich cores appear to follow a different transformation path in the solid state than the rest of the dendrites, substantially complicating the analysis of the phase sequencing and microstructure evolution during the post-solidification cooling. The scale of segregation can be refined by approximately one order of magnitude using melt-spinning or splat-quenching instead of arc melting. Homogenization has then been accomplished in a few hours.

CONSOLIDATION - HIP consolidation of titanium aluminide powders produced by Rapid Solidification Rate (RSR) and Rotating Electrode Process (REP) were studied (Choi et. al) and compared to computer model predictions (Ashby). In a corollary effort a high temperature sensor was developed (Paden et. al) for in-situ monitoring of the consolidation process. Characterization of the starting powders; RSR Ti-48%Al-2.5%Nb-0.3%Ta, and REP Ti-50%Al-2%Nb, showed dendrites with primarily hexagonal symmetry, indicative of α-phase formation from the liquid. There microstructures were consistent with solidification studies reported above - e.g. carbon contamination in the RSR process strongly favors the formation of primary α. Good agreement was obtained between the predicted HIP maps and the experimental data during later stages of densification (>85%) and longer times (4 and 16 hours). The power law creep mechanism predicted by the model is in best agreement with the experiments. However, discrepancies arise at the earlier stages. At lower temperatures, the model predicts much faster consolidation, indicating a required adjustment to the model. At higher temperatures and short times (1 hour), the model predicts a slower densification. Direct in-situ sensing, using an eddy current sensor, provided an explanation for the latter discrepancy. It was found that significant consolidation occurred during the pressurization process which the model does not take into consideration.

COMPOSITES - The composite processing facet of this research has explored using alloying additions that develop reinforcements in titanium aluminides, through suitable control of the chemistry and solidification parameters. The compounds of most interest are those which exhibit strongly anisotropic growth from the melt, either because of their crystal structure and/or interfacial tension characteristics. Plate reinforcements have been readily grown in the Ti-Al-C system, where the phase in equilibrium with γ-TiAl is Ti_2AlC, a hexagonal structure with a rather

12

complicated stacking sequence along the c-axis (Valencia et al.). Plates ~ 1 μm thick with aspect ratios of ~ 20 and up to 12% by volume have been grown as primary phases in Ti-(48–51at%)Al-3at%C. Significant increases (20–60%) in creep strength were observed in these alloys when compared with unalloyed γ, but there are indications that some of the strengthening is due to dissolved C in the matrix.

Borides appear more promising as reinforcing phases for γ-TiAl; they offer flexibility for microstructural control and relatively straightforward processing conditions. The structures of interest are all based on the trigonal prismatic coordination of 6 metal atoms around each B, which may be arranged in close-packed (TiB$_2$), columnar (TiB) or lamellar (TaB,NbB) arrays. The boride in equilibrium with the binary γ-TiAl is TiB$_2$, which forms a mixture of plates, needles and highly convoluted flakes when solidifying as a secondary phase. Major progress has been made in the understanding of these different morphologies and their dependence on alloy composition and solidification rate (Hyman et al.), but TiB$_2$ may not lend itself to the production of appreciable volume fractions of reinforcements with the desirable morphologies. However, controlled additions of Ta or Nb change the structure of the stable primary boride to TiB, with Ta or Nb partially substituting for Ti. These borides grow as rods ~ 1 to 5 μm in diameter with aspect ratios of 20 or larger, as expected from crystallographic considerations, and have been produced in volume fractions up to 15% (Valencia et al.). Current studies are focused on optimizing the scale and distribution of these reinforcements, as well as on exploring alloy compositions which may produce platelike phases based on the structure of TaB.

PLANAR FLOW CASTING - High-speed, high-magnification video techniques were used to investigate the geometry and stability of the puddle during planar flow casting of aluminum alloys. Puddle oscillations and surface instabilities are

13

correlated to macroscopic defects in the ribbon. The process was modelled numerically using fluid mechanics and heat transfer to establish thermal profiles in the solidifying puddle. Although not specifically reported herein, the system has been upgraded and used to produce ribbons of titanium aluminides for some of the solidification studies in this report.

CERAMICS

Pressureless densification of ceramic powders containing reinforcements has been emphasized as needed to understand processing limitations and to develop new methods for processing ceramic composites. Two major problems have been studied: the forming of powder compacts containing reinforcements and the constrained densification of the matrix powder. It has been shown that pressure filtration can be used to pack powder within a fiber preform provided: a) repulsive forces exist between particles and between the reinforcement and particles and b) the particle size to fiber diameter ratio is ≤ 0.01. The latter requirement not only provides for the flow of particles through preform channels, but also for high packing density within the preform. Densification of the powder matrix is constrained by the reinforcement. Studies of this problem have emphasized powders containing a non-connected network of inert particles, viz., Al_2O_3-SiC plates and Al_2O_3-ZrO_2 inclusions. Shrinkage measurements, mercury porosimetry, and microstructural observations illustrate the sequential events: sintering and shrinkage occurs until an interconnected dense network forms that surrounds lower density regions. Coarsening of grains within the lower density regions causes voids within these regions to grow and dissipate the sintering potential. *Higher particle packing density results in less differential shrinkage and higher composite density.*

Studies concerning sintering via vapor phase transport have been initiated with the goal of strengthening the powder matrix without the shrinkage associated

14

with other modes of material transport. Studies have also been initiated in which precursor ZrO_2 fibers are incorporated into powders which also shrink, either more or less than the powder, depending on their prior heat treatment.

MIXED MODE FRACTURE MECHANICS OF INTERFACES

John W. Hutchinson

MIXED MODE FRACTURE MECHANICS OF INTERFACES

John W. Hutchinson
Division of Applied Sciences
Harvard University
Cambridge, MA 02138

Introduction

This paper reviews some recent developments in the mechanics of fracture of interfaces between elastic materials. The review draws heavily on the recent work of the author and his collaborators at Harvard and the University of California, Santa Barbara.

Due to asymmetry in loading and elastic properties across the interface, many interfacial fracture problems are inherently mixed mode. In mixed mode, both normal and shear stresses act across the interface ahead of the tip of the interfacial crack, and both opening and shearing displacements occur on the crack faces behind the tip. Thus interfacial fracture in two-dimensional geometries involves mode 1 (opening) and mode 2 (shearing) stress intensity factors, and one must allow for a toughness characterization which, in general, is a function of the relative amounts of mode 1 and mode 2. This is one of the main differences between interfacial fracture mechanics and fracture mechanics for isotropic homogeneous materials in which mode 1 toughness receives predominant emphasis.

This review starts off with a discussion of universal crack tip fields at an interface between dissimilar isotropic elastic solids. Then, examples are given which illustrate the extent to which loading and moduli differences influence conditions at the tip. Some mixed mode toughness data are presented and related to preliminary micromechanics modeling. Decohesion of coatings and thin films is discussed within the context of interfacial fracture, and some first results of an analysis of the cut test are reported wherein decohesion is induced by a straight cut through a residually stressed coating. The final example deals with the deflection of a crack into an interface.

Crack Tip Fields

With the interface on the x_1-axis, let E_1, μ_1 and ν_1 be the Young's modulus, shear modulus and Poisson's ratio of material 1 lying above the interface ($x_2 > 0$) with similar quantities, E_2, μ_2 and ν_2, for material 2 lying below the interface. For plane problems with traction boundary conditions, only two nondimensional combinations of the four independent material moduli parameters enter into any solution (1). For plane strain the moduli mismatch parameters of Dundurs are

$$\alpha = \frac{\bar{E}_1 - \bar{E}_2}{\bar{E}_1 + \bar{E}_2} \quad \text{and} \quad 2\beta = \frac{\mu_1(1-2\nu_2) - \mu_2(1-2\nu_1)}{\mu_1(1-\nu_2) + \mu_2(1-\nu_1)} \tag{1}$$

where $\bar{E} = E/(1-\nu^2)$ is the plane strain tensile modulus. Note that α and β both vanish when dissimilarity between the elastic properties of the materials is absent, and α and β change signs when the materials are switched.

For each material pair, a universal singular crack tip field exists at the crack tip according to linear elasticity theory for a traction-free line crack. For the plane problems, the normal and shear stresses of the singular field acting on the interface a distance r ahead of the tip can be written in the compact "complex" form

$$\sigma_{22} + i\sigma_{12} = \frac{(K_1 + iK_2)r^{i\epsilon}}{\sqrt{2\pi r}} \tag{2}$$

where $i \equiv \sqrt{-1}$ and the oscillation index ϵ depends on β according to

$$\varepsilon = \frac{1}{2\pi} \ln\left(\frac{1-\beta}{1+\beta}\right) \qquad [3]$$

In plane strain the crack face displacements a distance r behind the tip, $\delta_i = u_i(-r,0^+) - u_i(-r,0^-)$, are given by

$$\delta_2 + i\delta_1 = \frac{4}{\sqrt{2\pi}} \frac{(1/\bar{E}_1 + 1/\bar{E}_2)}{(1+2i\varepsilon)\cosh(\pi\varepsilon)} (K_1 + iK_2)\sqrt{r}\ r^{i\varepsilon} \qquad [4]$$

The amplitude factors, K_1 and K_2, depend linearly on the applied loads and on the details of the full geometry of the body, as will be illustrated below. These stress intensity factors are defined to be consistent with corresponding stress intensity factors for cracks in homogeneous problems (2). The energy release rate per unit length of extension of the crack in the interface is related to the stress intensity factors by (for plane strain)

$$\mathcal{G} = \frac{(1/\bar{E}_1 + 1/\bar{E}_2)}{2\cosh^2\pi\varepsilon} (K_1^2 + K_2^2) \qquad [5]$$

which is the generalization of Irwin's famous result for a homogeneous isotropic material.

When $\varepsilon \neq 0$, the relative proportion of normal and shear stresses on the interface in the singular field varies slowly according to $r^{i\varepsilon} = \cos(\varepsilon \ln r) + i \sin(\varepsilon \ln r)$, and this feature complicates the implementation of interfacial mechanics in several respects. When $\varepsilon \neq 0$, the traction-free line crack solution is not fully consistent since the solution [4] implies that the crack faces interpenetrate behind the tip. Although seemingly troublesome, this inconsistency will usually be inconsequential since the distance of the contact region behind the tip is generally exceedingly small compared to the zone of nonlinear deformation or fracture processes. More problematic is the fact that K_1 and K_2 cannot be interpreted in a straightforward way as mode 1 and mode 2 intensity factors directly linked to normal and shear stresses. Moreover, the combination $K_1 + iK_2$ has dimensional units of stress σ and length L in the form $\sigma\sqrt{L} L^{-i\varepsilon}$. A peculiar consequence of the dimensional form is that the relative proportion of a critical combination (K_1, K_2) changes when units are changed (2).

The features mentioned above have stood as conceptual stumbling blocks in the development of interfacial fracture mechanics even though there is no compelling evidence to date to suggest that "ε-effects" are important. For many pairs of materials of interest one finds that ε is very small, e.g. less than a few hundredths in magnitude, even when α is substantial corresponding to ratios of the $\bar{E}$'s of 4 or 5 or more. An extensive tabulation of (α,β) pairs is presented in (3). At this stage in the development of interfacial fracture mechanics it is likely that the role of ε is far less important than other problems and issues, including even the simplification of elastic isotropy. Thus, various proposals have been put forth for ignoring ε. A consistent approach proposed in (4,5) is to systematically take $\beta = 0$ (and thus $\varepsilon = 0$), both in the determination of critical toughness data from test specimens and in subsequent application of such data to predict fracture. In other words, for a given material pair one contemplates another pair with slightly perturbed elastic moduli leading to $\beta = 0$ for the purpose of analyzing specimens and making applications.

With $\varepsilon = 0$, $r^{i\varepsilon} = 1$ in [2] and [4], and then K_1 and K_2 have conventional interpretations as mode 1 and mode 2 stress intensity factors measuring the singularity of normal and shear stresses, respectively, on the interface ahead of the tip. At various points throughout this review the assumption is that $\varepsilon = 0$ will be invoked in the above spirit. Recent work (6-9) has dealt with crack tip fields for an interface crack lying on an interface between dissimilar anisotropic elastic solids, and general restrictions on the moduli leading to nonoscillatory fields and conventional stress intensity factors have been identified.

Some Basic Solutions

A number of relatively simple exact solutions were given in the original papers on interface cracks from the 1960's (see (2) for a listing of these references). For a crack of length 2a lying on the interface between two semi-infinite blocks subject to remote stress (see Fig. 1a), the stress intensity factors of the right-hand tip are

$$K_1 + iK_2 = (\sigma_{22}^{\infty} + i\sigma_{12}^{\infty})(1+2i\epsilon)(2a)^{-i\epsilon}\sqrt{\pi a} \tag{6}$$

This solution does not depend on α, and perhaps it is partly for this reason that so much emphasis was placed on the role of ϵ in the beginning development of the subject. When $\epsilon = 0$ (recalling that the magnitude of ϵ is often only a few hundredths or less), [6] reduces to the classical result for the mode I and mode II stress intensity factors for a crack in a homogeneous, isotropic elastic solid.

Next, consider the semi-infinite interface crack between two infinite elastic layers shown in Fig. 1b and subject to the general loading indicated. This problem has been solved completely in (10). An exact expression for the energy release rate $\mathcal{G}$ can be obtained in closed form, which by [5] gives the magnitude of $K_1 + iK_2$. Integral equation methods have been used to obtain the relative proportion of K_2 to K_1. Complete details of the solution are given in (10); the form of the solution is

$$K_1 + iK_2 = \left[c_1 P h^{-1/2} + c_2 M h^{-3/2} \right] h^{-i\epsilon} \tag{7}$$

where P (load per unit thickness) and M (moment per unit thickness) are linear combinations of the P_i's and M_i's, and c_1 and c_2 are dimensionless complex numbers depending on α, β and h/H. Application of this general result will be made to two problems in the next section.

Another complete solution is available for sandwich geometries wherein a thin layer of material 2 is sandwiched between two planar bodies of material 1 with an interface crack lying between the thin layer and the upper block. Let K_I and K_{II} be the conventional mode I and mode II stress intensity factors for the homogeneous body of material 1 in the absence of the middle layer. If the thickness of the sandwich layer h is very small compared to the crack length and to all other relevant in-plane length quantities, a universal asymptotic relation exists between the interface intensity factors, K_1 and K_2, and the stress intensity factors, K_I and K_{II}, for the homogeneous problem. The factors K_I and K_{II}, can be thought of as applied, or remote, stress intensities for the semi-infinite crack problem depicted in Fig. 1c. The asymptotic relation between the interface stress intensities and the applied stress intensities is

$$K_1 + iK_2 = \sqrt{\frac{1-\alpha}{1-\beta^2}} \ (K_I + iK_{II}) h^{-i\epsilon} e^{i\omega(\alpha,\beta)} \tag{8}$$

where the real function $\omega(\alpha,\beta)$ is tabulated in (5). Thus, for example, if stress intensity factors K_I and K_{II} are known for a test specimen with homogeneous moduli, [8] is the universal relation giving the interface intensity factors for the corresponding sandwich specimens. The energy release rate $\mathcal{G}$ in this asymptotic problem is equal to the applied energy release rate $(1-\nu_1^2)(K_I^2 + K_{II}^2)/E_1$, as is readily seen from [8] with [5]. The β-dependence of ω is weak; $\omega(\alpha,0)$ is plotted in Fig. 2. For $\beta = 0$, [8] reduces to

$$K_1 + iK_2 = \sqrt{1-\alpha} \ (K_I + iK_{II}) e^{i\omega(\alpha,0)} \tag{9}$$

so that ω represents the shift in "phase" of the interface intensity factors relative to the applied intensity factors, i.e.

$$\tan^{-1}(K_2/K_1) = \tan^{-1}(K_{II}/K_I) + \omega \tag{10}$$

From Fig. 2 it is seen that this shift, which is due exclusively to the moduli mismatch, ranges between $5°$ and $-15°$, depending on α. For most material pairs this will not be a large effect.

The examples shown in Fig. 3 have been chosen to illustrate the range of mixed mode conditions which can arise in applications of interfacial fracture mechanics. For discussion purposes (as well as in actual implementation in many cases), it will be assumed that $\beta = 0$. Let

$$\psi = \tan^{-1}(K_2/K_1) \tag{11}$$

measure the "phase" of the stress intensity factors so that $\psi = 0°$ corresponds to pure mode 1 and $\psi = 90°$ corresponds to pure mode 2.

The *double cantilever specimen* in Fig. 3a is a special case of the two-layer problem in Fig. 1b. Specializing the results of (10) to this specific loading and geometry gives (with $\beta = 0$)

$$K_1 + iK_2 = 2\sqrt{3} \ Mh^{-3/2}e^{i\psi} \tag{12}$$

where ψ is plotted as a function of α in Fig. 4. The symmetry of the geometry and the loading dictates that the specimen must be pure mode 1 when there is continuity of moduli across the interface (i.e. when $\alpha = 0$). When $\alpha \ne 0$, the asymmetry due to the moduli mismatch induces a mode 2 component which can be appreciable when, for example, $\alpha = \pm.5$, corresponding to $E_1/E_2 = 3$ or $1/3$.

Figure 3b depicts a *thin film or coating* of material 1 with residual biaxial tensile stress σ attached to a thick substrate of material 2. The film is decohering along the interface driven by the residual stress. When the decohesion crack length is long compared to the film thickness, conditions at the tip approach a steady state which is independent of crack length. For the limit when the film thickness h is very small compared to the substrate thickness H (i.e. $h/H \rightarrow 0$), the steady state energy release rate is

$$\mathbf{S} = \frac{1}{2}(1-v_1^2)\sigma^2 h/E_1 \tag{13}$$

The interface intensity factors can be obtained by specializing the results in (10). When $\beta = 0$,

$$K_1 + iK_2 = \sqrt{\frac{1-\alpha}{2}} \ \sigma h^{1/2}e^{i\psi} \tag{14}$$

where ψ is plotted in Fig. 5. It is noted that decohesion cracking is inherently mixed mode with ψ ranging from 45° to 70°, depending on α.

The *four-point bend specimen* in Fig. 3c has been analyzed and used for measuring interfacial toughness for several material combinations (11, 12). The specimen has the advantage that the crack tip is in steady-state condition when the crack is long compared to the thickness but still between the inner loading points. The results for the two-layer problem in Fig. 1b can be used to analyze the specimen when the crack is advancing under steady-state conditions. The analysis, which is given in (10), is highly accurate when the total thickness of the specimen is small compared to the distance between the loading points. Over a fairly wide range of α and h/H the analysis predicts $\psi \cong 45°$. Thus, this specimen measures mixed mode toughness under conditions when K_1 and K_2 are approximately equal.

The final example considered is the *fiber pull-out problem* depicted in Fig. 3d. This problem is of considerable interest in connection with the role of fibers in bridging matrix cracks in brittle matrix composites. For the fiber to perform effectively it is essential that it debond from the matrix some distance away from the main matrix crack. This debonding permits fiber pull-out and a consequent toughening contribution from the fiber. There are a variety of conditions under which it is important to understand the fiber debonding process, and many

of these remain to be analyzed. One numerical analysis (13), applicable when there is a residual tensile stress acting across the fiber/matrix interface prior to debonding, predicts that the interface debonding crack is heavily mode 2, as might be expected from the loading and geometry, with $\psi \cong 75°$.

Interfacial Toughness

The examples of Fig. 3 are intended to drive home the point that interfacial fracture conditions can range from mode 1 to mode 2 and are often decidedly mixed. The specification of interface toughness when $\varepsilon \neq 0$ has been discussed in (2). As noted earlier, the specification is simplified when $\varepsilon = 0$ since K_1 and K_2 then have straightforward interpretations as mode 1 and mode 2 intensity factors. Here we will confine the discussion to material pairs for which $\varepsilon = 0$, either exactly or by approximation.

In general one must expect that the toughness of the interface (i.e. the critical value of $\mathfrak{S}$ associated with crack advance in the interface) is a function of the combination of K_1 and K_2 governing the crack tip stresses. That is, the specification of interface toughness requires that $\mathfrak{S}_c$ be known as a function of $\psi = \tan^{-1}(K_2/K_1)$. The attainment of conditions for crack advance in the interface is stated formally as

$$\mathfrak{S} = \mathfrak{S}_c(\psi) \tag{15}$$

where $\mathfrak{S}$ is given in terms of K_1 and K_2 by (5). It may be that an "ideally brittle" interface will have a toughness $\mathfrak{S}_c$ which is independent of ψ and equal to the energy needed to create the two free surfaces. No mixed mode toughness data exists yet to show that "ideally brittle" interfaces actually exist. In fact, very little experimental toughness data exists which systematically spans a range of ψ. One recent set of data for $\mathfrak{S}_c(\psi)$, shown in Fig. 6, was generated in (12) for an epoxy bonded to glass. The experimental data was obtained using a variety of specimens such as those in Fig. 3 to span the range of ψ from pure mode 1 to $\psi \cong 75°$. For this material pair and for the conditions under which the surfaces were bonded, the toughness is clearly a strong function of the combination of modes at the crack tip.

A first attempt at a micromechanics model of interfacial toughness as it is influenced by non-planarity of the fractured interface was given in (14). The idealized situation envisioned is depicted in Fig. 7a. Due either to the intrinsic faceting of the interface in the bonded state (15) or to roughness generated by the fracture process itself, the fracture surfaces are taken to have facets of height w and spacing l. Within the vicinity of the crack tip where the crack opening displacement, δ_2, is less than w, the facets on opposite faces contact one another and tend to shield the tip from the full effect of K_2 when K_2 is not zero. The analysis in (14), which neglects any dissimilarity in moduli, derives an estimate for $\mathfrak{S}_c(\psi)$ based on the criterion that the shielded energy release rate at the tip equals the "intrinsic" toughness $\mathfrak{S}_c^0$, which can also be identified with the pure mode 1 toughness.

The effective interface toughness depends on a single nondimensional material/roughness parameter R which is predominantly dependent on

$$\frac{Ew^2}{l\mathfrak{S}_c^0} \tag{16}$$

(A full specification of R is given in (14).) Plots of the normalized toughness increase as a function of ψ for different levels of interface roughness R are shown in Fig. 7b. Included in Fig. 7b is the experimental data of Fig. 6 from (12) for the epoxy/glass interface. The measured roughness of the fractured interface is in approximate agreement with the roughness level $R \cong 1$ indicated by the model predictions in Fig. 7b (12). In the limiting result for $\mathfrak{S}_c(\psi)$ when $R \to \infty$, K_2 has no influence on the energy release rate at the tip, as the model in Fig. 7a would suggest. Thus, for all ψ the energy release rate at the tip is $(1-v^2)K_1^2/E$ and, consequently,

$$\mathfrak{S}_c(\psi) = \mathfrak{S}_c^0[1 + \tan^2\psi] \tag{17}$$

Clearly the epoxy/glass interface behaves more like a highly rough interface than an ideally brittle one.

Other microscopic processes can be expected to influence the mixed-mode toughness. These include friction and plasticity, either dislocation emission at the tip or more extensive zones of high dislocation density.

Decohesion of Thin Films

A necessary condition for a thin film under residual biaxial tension σ to undergo extensive decohesion can be inferred from the results [13] and [14] for the plane strain problem in Fig. 3b. When the substrate is thick compared to the film ($h/H \to 0$) the critical stress for steady-state, two-dimensional cracking is

$$\sigma_c = \sqrt{\frac{2 \mathfrak{I}_c(\psi)E_1}{(1-v_1^2)h}} \qquad [18]$$

where ψ is obtained from Fig. 5. Decohesion of regions whose extent is large compared to the thickness will only occur if $\sigma > \sigma_c$. In the *cut test* a straight cut, many times h in length, is made through the film. If $\sigma > \sigma_c$, a region of decohesion spreads from the cut. The equilibrium shape of the decohered region is governed by the condition that the crack driving force, $\mathfrak{I}$, at every point along the edge of the decohesion just attains the condition for further decohesion.

Because both normal and shear stresses act along the decohesion boundary, the interface crack front is fully mixed mode with mode 3 present as well as modes 1 and 2. The energy release rate under the three modes is (with $\beta = 0$)

$$\mathfrak{I} = \frac{1}{2}\left(\frac{1}{E_1} + \frac{1}{E_2}\right)\left(K_1^2 + K_2^2\right) + \frac{1}{4}\left(\frac{1}{\mu_1} + \frac{1}{\mu_2}\right)K_3^2 \qquad [19]$$

Along the decohesion boundary in the cut test the nature of the local stressing is such that the relative proportion of K_2 to K_1 is fixed with $\psi = \tan^{-1}(K_2/K_1)$ given by Fig. 5. Thus the shape of the decohered region can not be used to discriminate the role of K_2 in the decohesion criterion. However, the shape of the boundary is sensitive to how K_3 influences the decohesion criterion. An extensive study of this shape dependence is given in (16). Here we will report the shapes of the decohesion boundary for two limiting decohesion criteria suggested in the previous section: (i) based on an ideally brittle interface with $\mathfrak{I} = \mathfrak{I}_c^0$ where $\mathfrak{I}_c^0$ is mode-independent, and (ii) based on a critical value of K_1, independent of K_2 and K_3.

The shapes of the decohesion boundary based on the criterion $\mathfrak{I} = \mathfrak{I}_c^0$ are shown in Fig. 8a. The critical stress parameter σ_c is given by [18] with $\mathfrak{I}_c(\psi) = \mathfrak{I}_c^0$. Once the decohesion criterion has been specified, the shape and extent of the decohesion depends only on σ/σ_c and the Poisson's ratio of the film v_1. The shapes corresponding to the criterion $K_1 = K_1^c$, independent of K_2 and K_3, are shown in Fig. 8b. The critical stress is still given by [18], but now

$$\mathfrak{I}_c(\psi) = \frac{1}{2}\left(\frac{1}{E_1} + \frac{1}{E_2}\right)\left(K_1^c\right)^2\frac{1}{\cos^2\psi} \qquad [20]$$

where ψ is obtained from Fig. 5. The most notable difference between the two families of shapes is that the decohesion boundary stands off the end of the cut when the criterion $\mathfrak{I} = \mathfrak{I}_c^0$ holds but remains attached when the criterion is based on a critical value of K_1. Observation of the cut test for a polyamide/glass system indicates

that the decohesion boundary remains attached for this systems (16). Furthermore, by considering a range of criteria, involving varying proportions of the mode 3 contribution, and the corresponding shape dependence, it is possible to draw some conclusions about the quantitative role of mode 3 on the decohesion criterion (16).

Crack Deflection Versus Penetration at an Interface

Whether an interface crack propagates within the interface or kinks out of the interface into one of the adjoining materials depends in an essential way on the combination of modes acting at the tip of the crack (4). It is also true that the tendency of a crack approaching an interface to deflect into the interface, as opposed to penetrating through the interface, depends essentially on mixed mode considerations, as will now be illustrated using an example from (17).

Consider the crack in Fig. 9a impinging at a right angle to the interface and assume the loading is symmetric with respect to the crack. Now consider two competing trajectories of crack advance: penetration of the interface in Fig. 9b, and deflection into the interface in Fig. 9c. In each case, the energy release rate, $\mathcal{G}$, is proportional to $a^{1-2\lambda}$, where a is the putative crack length and λ is the exponent of the stress singularity of $(\sigma \sim r^{-\lambda})$ of the problem in Fig. 9a. The ratio, $\mathcal{G}_d/\mathcal{G}_p$, is independent of a when a is small compared to the length of the main crack. Thus, the *relative* tendency of the crack to deflect or penetrate can be assessed using this ratio. Figure 10a shows this ratio as a function of α when $\beta = 0$. The proportion of mode 2 to mode 1 for the deflected crack is shown in Fig. 10b. (The penetrating crack is in pure mode 1.)

The condition (a necessary condition) for the crack to be deflected into the interface, rather than to penetrate it, is

$$\frac{\mathcal{G}_c(\psi)}{\mathcal{G}_c^{(1)}} < \frac{\mathcal{G}_d}{\mathcal{G}_p} \tag{21}$$

where $\mathcal{G}_c(\psi)$ is the toughness of the interface at the relevant combination of modes and $\mathcal{G}_c^{(1)}$ is the mode 1 toughness of material 1 across the interface. When the elastic mismatch is not large ($\alpha \equiv 0$), the condition requires that the interface toughness be less than about 1/4 the toughness of material 1, where the interface toughness is that associated with $\psi \equiv 45°$. Deflection as a symmetric doubly-deflected crack has also been considered in (17). Generally, the singly-deflected crack of Fig. 9c is the controlling case, but this conclusion depends somewhat on how strongly $\mathcal{G}_c(\psi)$ depends on ψ since the ψ-values differ for the two cases.

Acknowledgements

References

1.   J. Dundurs, J. Appl. Mech., 36, p. 650 (1969)
2.   J. R. Rice, J. Appl. Mech., 55, p. 98 (1988).
3.   T. Suga, G. Elssner and S. Schmauder, J. Composite Materials, p. 917 (1988).
4.   M-Y. He and J. W. Hutchinson, "Kinking of a crack out of an interface", to be published in J. Appl. Mech.
5.   Z. Suo and J. W. Hutchinson, "On sandwich test specimens for measuring interface crack toughness", to be published in Materials Science and Engineering.
6.   J. L. Bassani and J. Qu, "Finite crack on bimaterial and bicrystal interfaces", to be published in J. Mech. Phys. Solids.

7   T. C. T. Ting, Int. J. Solids Struct., 22, p. 965 (1986).

8.   V. K. Tewary, R. H. Wagoner and J. P. Hirth, J. Mater. Res. 4, p. 124 (1989).

9.   Z. Suo, "Singularities, interfaces and cracks in dissimilar anisotropic media", to be published.

10. Z. Suo and J. W. Hutchinson, "Interface crack between two elastic layers", to be published in Int. J. of Fracture.

11. P. G. Charalambides, J. Lund and A. G. Evans, "A test specimen for determining the fracture resistance of bimaterial interfaces", to be published in J. Appl. Mech.

12. H. C. Cao and A. G. Evans, "An experimental study of the fracture resistance of bimaterial interfaces", to be published in Mech. of Maters.

13. P. G. Charalambides and A. G. Evans, "the interfacial fracture resistance in brittle matrix composites" to be published in J. Am. Ceram. Soc.

14. A. G. Evans and J. W. Hutchinson, "Effects of non-planarity of the mixed mode fracture resistance of bimaterial interfaces", to be published in Acta metall.

15. W. Mader and M. Rühle, "Electron microscopy studies of defects at diffusion-bonded Nb/Al_2O_3 interfaces", to be published in Acta metall.

16. H. Jensen, J. W. Hutchinson and K. S. Kim, work in progress on decohesion in the cut test.

17. M-Y. He and J. W. Hutchinson, "Crack deflection at an interface between dissimilar elastic materials", to be published in Int. J. Solids Struct.
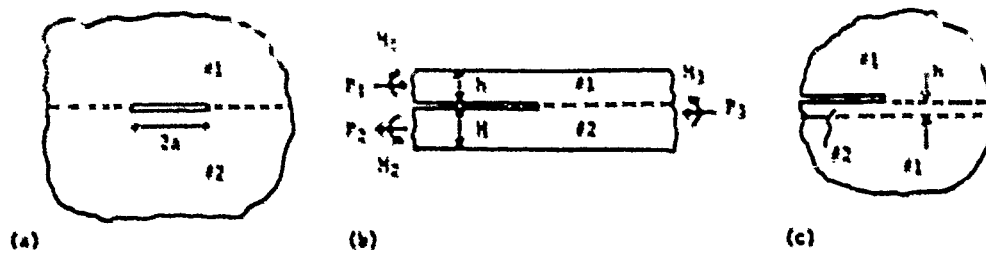
Fig. 1 Three basic interface crack geometries.
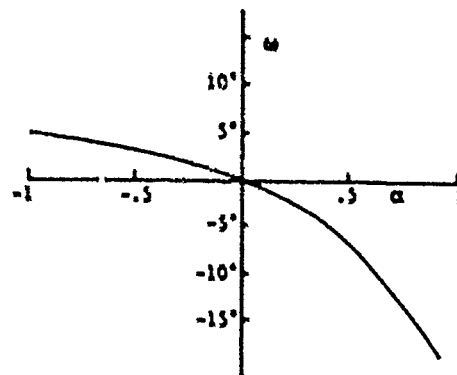
Fig. 2 Phase shift ω between remote and near-tip stress intensity factors $(\beta = 0)$.
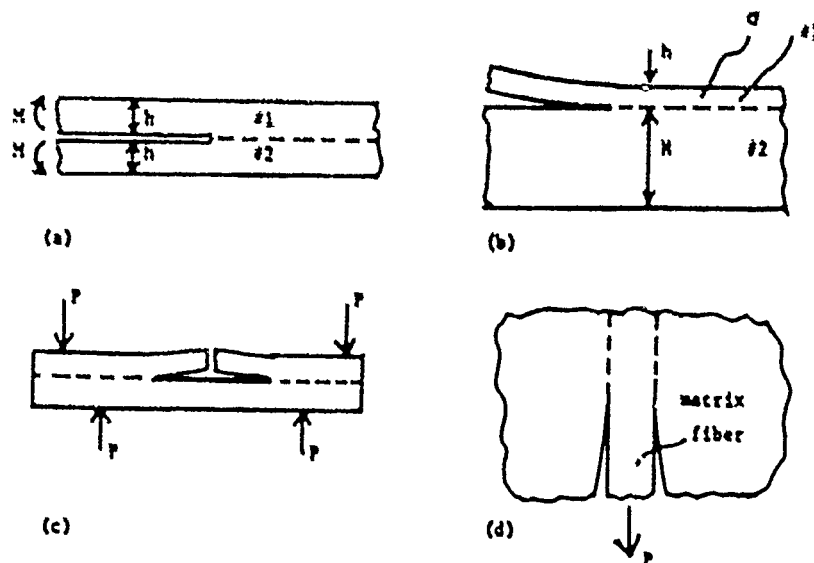
Fig. 3 Four examples illustrating range of mixed-mode conditions at the tip of an interface crack: (a) double cantilever, (b) decohesion of thin film under residual biaxial tensile stress, (c) four-point bend specimen, and (d) fiber pull-out.
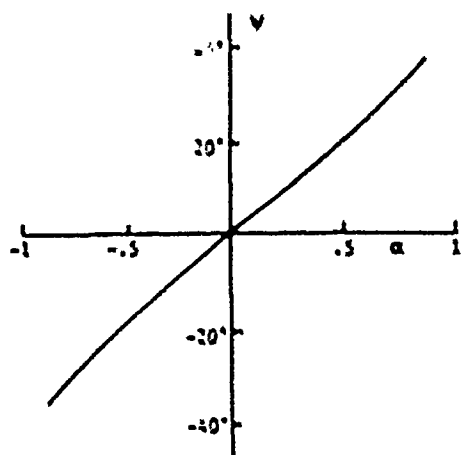
Fig. 4 Phase of stress intensities, $\psi = \tan^{-1}(K_2/K_1)$, for double cantilever in Fig. 3a ($\beta=0$).
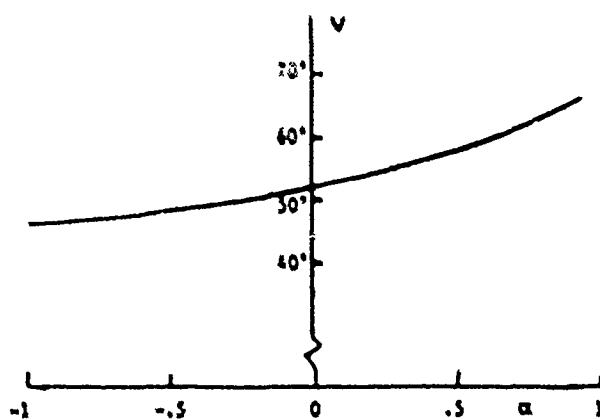
Fig. 5 Phase of stress intensities, $\psi = \tan^{-1}(K_2/K_1)$, for decohesion of thin film in Fig. 3b ($\beta=0$).
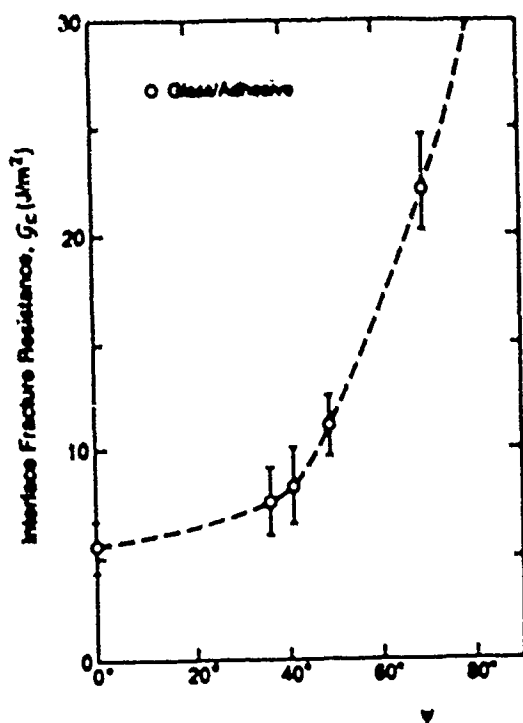
Fig. 6 Interface toughness data for an epoxy/glass system from (12).
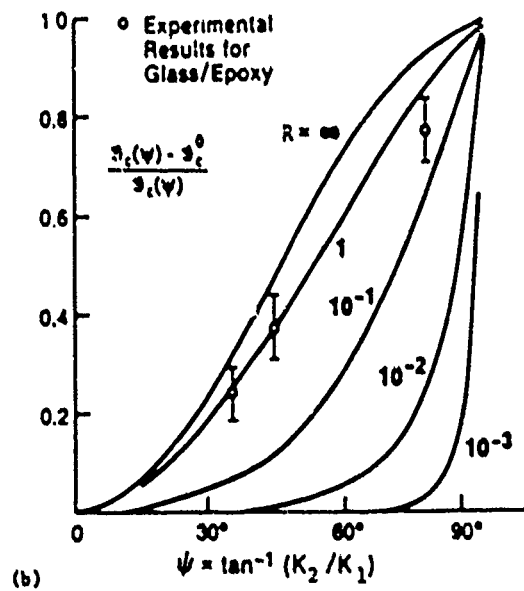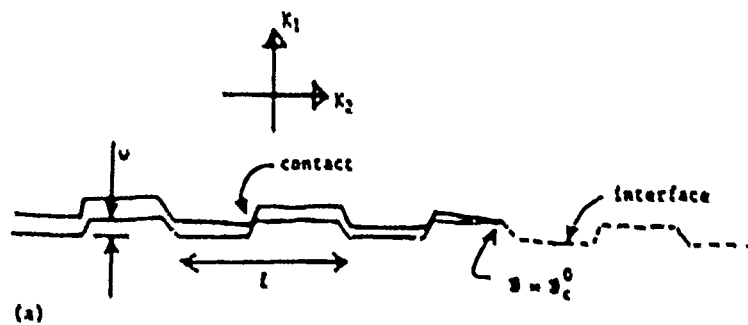
Fig. 7   Micro-mechanical model of effect of interface roughr   ¬ixed-mode toughness.
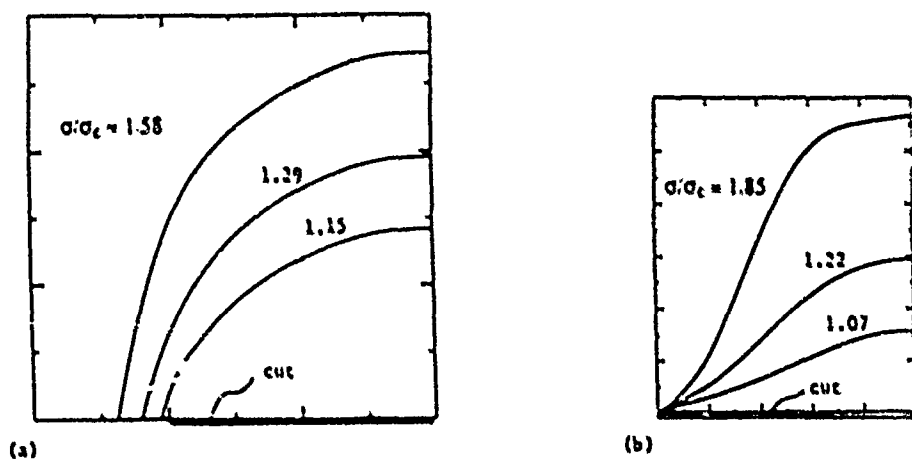
Fig. 8 Shapes of decohesion region in the cut test at various levels of σ/σ_c ($v_1 = 1/3$): (a) based on criterion for an ideally brittle interface, $S = S_c^0$; (b) based on the criterion $K_1 = K_1^c$ independent of K_2 and K_3.
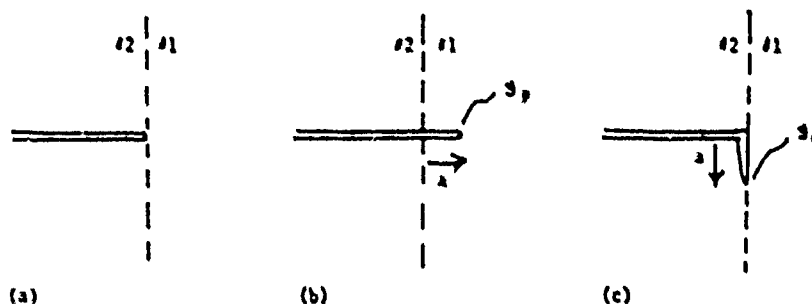
Fig. 9 Competition between crack penetration and deflection at an interface: (a) main crack at interface, (b) penetration of interface, and (c) deflection by interface.
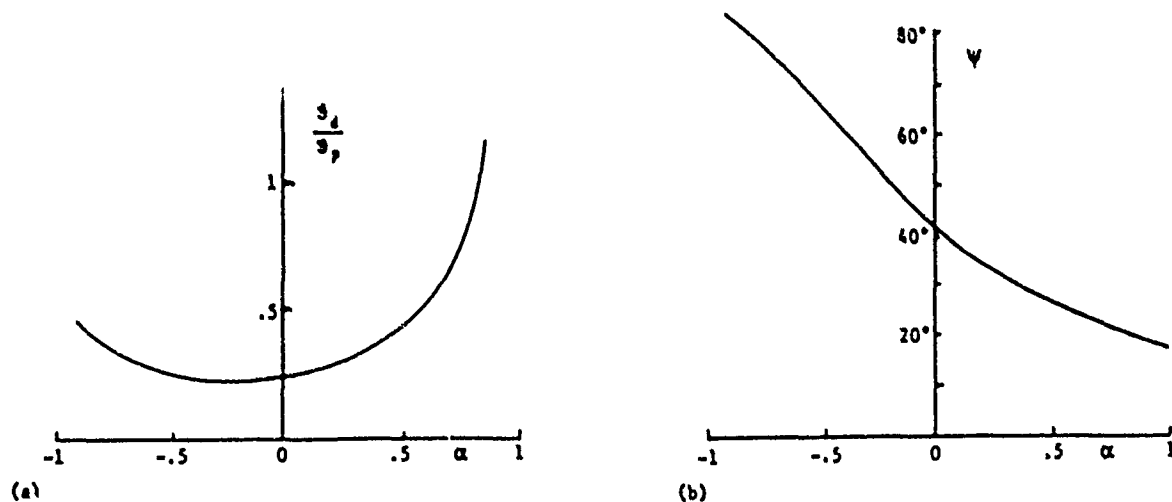
Fig. 10 (a) Ratio of energy release rate of deflected crack to that of penetrating crack as a function of the elastic mismatch parameter α ($\beta=0$). (b) Phase of stress intensities, $\psi = \tan^{-1}(K_2/K_1)$, for deflected crack tip ($\beta=0$).

MATERIALS
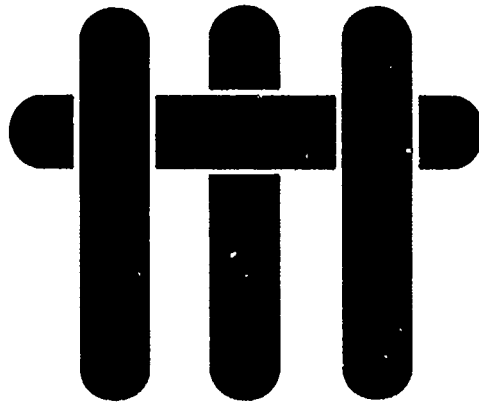
A METHOD FOR CALCULATING STRESS INTENSITIES IN BIMATERIAL FRACTURE

P. P. L. Matos, R. M. McMeeking, P. G. Charalambides
and M. D. Drory

Materials and Mechanical Engineering Departments
College of Engineering
University of California, Santa Barbara
Santa Barbara, California 93106

A method for calculating stress intensities in bimaterial fracture

P.P.L. MATOS, R.M. McMEEKING, P.G. CHARALAMBIDES and M.D. DRORY
Department of Materials and Department of Mechanical Engineering, University of California, Santa Barbara, California 93106, USA

Abstract. A numerical method is presented for obtaining the values of K_I^*, K_{II}^* and K_{III}^* in the elasticity solution at the tip of an interface crack in general states of stress. The basis of the method is an evaluation of the J-integral by the virtual crack extension method. Individual stress intensities can then be obtained from further calculations of J perturbed by small increments of the stress intensity factors. The calculations are carried out by the finite element method but minimal extra computations are required compared to those for the boundary value problem. Very accurate results are presented for a crack in the bimaterial interface and compared with other methods of evaluating the stress intensity factors. In particular, a comparison is made with stress intensity factors obtained by computing J by the virtual crack extension method but separating the modes by using the ratio of displacements on the crack surface. Both techniques work well with fine finite element meshes but the results suggest that the method that relies entirely on J-integral evaluations can be used to give reliable results for coarse meshes.

1. Introduction

In mixed mode elastic fracture mechanics it is desirable to evaluate accurately the individual mode I, II and III crack tip stress intensity factors K_I, K_{II} and K_{III} [1]. The modes are tensile opening (I), in plane shear (II) and antiplane shear (III) where the relevant plane is such that the crack front passes through it orthogonally. In addition, when a crack lies with the near tip segment in the interface between two linear elastic isotropic materials, the parameters K_I^*, K_{II}^* and K_{III}^* can be used to characterize the near tip field [2–4]. With the coordinate system aligned as in Fig. 1, the inplane stresses ahead of the crack on $x_2 = 0$ are given asymptotically by

$$\sigma_{22} + i\sigma_{12} = \frac{(K_I^* + iK_{II}^*)x_1^\varepsilon}{\sqrt{2\pi x_1}},$$

(1)

where

$$\varepsilon = \frac{1}{2\pi} \log\left[\left(\frac{\mu_1}{G_1} + \frac{1}{G_2}\right)\Big/\left(\frac{\mu_2}{G_2} + \frac{1}{G_1}\right)\right]$$

(2)

and $i = \sqrt{-1}$. The parameters for the materials are such that G_i are the shear moduli, E_i are the Young's moduli and ν_i are the Poisson's ratios. The parameters μ_i are defined so that they equal $(3 - \nu_i)/(1 + \nu_i)$ in plane stress and otherwise are $3-4\nu_i$. When the two materials are identical $\varepsilon = 0$ and the factors K_I^* and K_{II}^* become identical to K_I and K_{II} respectively. The antiplane stress field for the isotropic bimaterial case is unchanged from the homogeneous case, but the stress intensity factor is denoted K_{III}^* for the former for clarity.
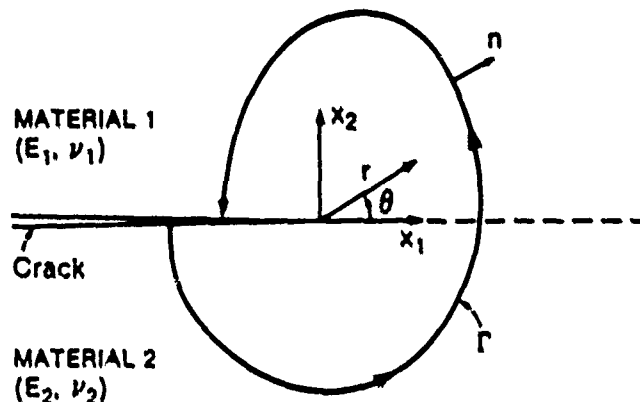
Fig. 1 Coordinates and typical contour used to evaluate the J-integral.

Henceforth, the discussion and development will be confined to the inplane stresses and displacements and the antiplane component of the near tip fields will be neglected. However, the methods addressed in this paper have obvious generalizations when there is antiplane stressing.

It is important to note that the factors K_I^* and K_{II}^* do not represent opening and shear modes [2–4]. Instead, the oscillatory nature of (1) and the associated displacements mean that the ratio of modes in the crack surface motions varies with distance from the crack tip. However, the wavelength of the oscillations increases with distance from the crack tip.

When the material interface is a plane, the J-integral [5] is path independent for the crack segment including the tip that lies in the material boundary. The relationship between J and the stress intensity factors is given by [6]

$$J = \frac{1}{H}(K_I^{*2} + K_{II}^{*2}).$$ (3a)

where

$$\frac{1}{H} = \frac{1}{2}\left(\frac{1}{E_1'} + \frac{1}{E_2'}\right) \Big/ \cosh^2(\pi\epsilon)$$ (3b)

and $E_i' = E_i$ in plane stress and $E_i' = E_i/(1 - v_i^2)$ in plane strain.

A review of numerical methods for obtaining the single mode stress intensity factors in the homogeneous case has been given by Gallagher [7]. The most effective of these is the virtual crack extension method of Parks [8]. However, as in all methods which evaluate J or the energy release rate, the technique must be augmented if it is to be used to establish the individual stress intensities in the mixed mode case. It is possible to define associated path independent integrals and this has been done by Ishikawa, Kitigawa and Okumura [9] for the homogeneous case. The technique was extended by Bui [10] and consists of taking the parts, symmetric and antisymmetric about the crack plane, of the planar displacement, strain and stress fields and using them separately in path integrals. One integral can be manipulated to give K_I and the other to provide K_{II}. As noted by Bui [10], the virtual crack extension

method can be used to evaluate the path integrals. However, the approach of Ishikawa et al. [9] and Bui [10] cannot be used for cracks located in bimaterial interfaces. This follows from the fact that the displacements solution in one material does not satisfy the governing equations in the other, and so the symmetric and antisymmetric parts are invalid displacements in both materials.

Yau, Wang and Corten [11] following Stern, Becker and Dunham [12] developed a technique for obtaining the separate modes by using the M_1 integral of Chen and Shield [13]. When two solutions for a given geometry are added together, a third field results for which the J-integral can be found. The value of this J differs by M_1 from the sum of the J values for the individual solutions. Thus, when one of the solutions is of a single mode of known magnitude, the value of M_1 can be used to determine one of the stress intensity factors of the other solution. Wang et al. [11] used this method for problems in homogeneous materials. They obtained the required numerical results by the finite element technique and evaluated the J and M_1 integrals by taking paths through the finite element mesh. Accurate results were obtained.

It is clear that the method of Wang et al. [11] can be used for a crack in bimaterial interface, since J and M_1 integrals still pertain. Indeed, Wang and Yau [14] applied the technique to the bimaterial case and Rice [15] has stated a line integral equivalent to M_1 specifically for the bimaterial case. Furthermore, since M_1 has J-like character, it can be evaluated by the virtual crack extension method [8]. This is desirable, since highly accurate results can be obtained at low cost in this way. Indeed Shih and Asaro [16] have recently used an equivalent method to obtain K_1^* and K_{11}^* from finite element solutions.

In this paper, a virtual crack extension based method for obtaining K_1^* and K_{11}^* from finite element solutions will be described. Results for a few problems will be presented and compared to known accurate solutions. The dependence of the accuracy on the refinement of the mesh will be investigated. Finally, results for K_1^* and K_{11}^* obtained by using crack surface displacements from finite element calculations for some problems will be compared to values resulting from the new method.

2. Formulation

The J-integral [5] is defined as

$$J = \int_\Gamma W \, dx_2 - n_i \sigma_{ij} \frac{\partial u_j}{\partial x_1} \, ds,$$ (4)

where Γ is any contour from the bottom crack surface around the tip to the top surface. In Fig. 1, n is the outward unit normal to the contour. W is the strain energy density, u are the displacements and ds is an infinitesimal element of contour arc length. The integral is path independent when the crack is straight, traction free and any material interface is parallel to the crack. If there are 2 displacement fields u^i and u^{ii} representing 2 solutions to 2 different boundary value problems for the crack, the values of J associated with them are J_i and J_{ii}. When the displacement fields are summed to give u^c, the value of J resulting is [13].

$$J_c = J_i + J_{ii} + M_1.$$ (5)

where M_1 is an integral given by

$$M_1 = \int_\Gamma \sigma_{ij}' \epsilon_{ij}^s \, dx_2 - \left(n_i \sigma_{ij}^s \frac{\partial u_j'}{\partial x_1} + n_i \sigma_{ij}' \frac{\partial u_j^s}{\partial x_1} \right) dS, \tag{6}$$

which is also path independent. From (3a), (3b) and (5) it can be deduced that

$$M_1 = \frac{2}{H} (K_1^* K_{1a}^* + K_{11}^* K_{11a}^*). \tag{7}$$

which was stated by Chen and Shield [13] in the homogeneous case. They also suggested that by use of a known field with $K_{11a}^* = 0$, the integral could be utilized to evaluate K_{1a} for the other field, a technique taken up by Yau et al. [11]. Later, Wang and Yau [14] used known displacement fields for cracks in bimaterial interfaces along with the M_1 integral to evaluate parameters of relevance to the crack tip fields in other solutions for bimaterial cracks. However, there is an ambiguity in their results of a type addressed recently by Rice [17] relating to dimensionality and phase of the quantity $K^* = K_1^* + iK_{11}^*$.

Following Parks [8], the energy release rate definition of J can be used to develop numerical methods for its evaluation. As a result

$$J = -(\partial U/\partial a)_f \approx -\tfrac{1}{2}\{u_n\}^T (\partial \{S\}/\partial a) \{u_n\}. \tag{8}$$

where U is the potential energy of the body and the differentiation with respect to crack length a is carried out at fixed load. The vector $\{u_n\}$ contains as elements the nodal degrees of freedom for a finite element calculation and $\{S\}$ is the stiffness matrix for the mesh of elements used to solve the crack problem. In the virtual crack extension method [8] the crack problem is first solved to obtain the vector $\{u_n\}$. Then a small virtual crack extension is caused in the plane of the crack and a new value of the stiffness computed. Usually the crack is extended by rigidly moving a core of elements around the tip and distorting only one ring such as the shaded one shown in Fig. 2. That is, all elements outside the distorted ring are also held rigid. Consequently in the core of rigid elements and in the outer rigid area, $\partial \{S\}/\partial a = 0$. As a result, the computations of changed stiffness required are limited and the multiplication involved in (8) involves small vectors and matrices. The method has been used extensively to compute single mode stress intensity factors in the homogeneous case.

Now consider a bimaterial problem A for which it is desired to compute K_1^* and K_{11}^*. First, solve the problem by the finite element method to find $\{u_n\}$. From this, compute J by the virtual crack extension technique summarized by (8). Then add to $\{u_n\}$ the displacements $\{\Delta u_n\}_i$ for a problem in the same geometry for which $K_{11}^* = 0$ and $K_1^* = \Delta K_1^*$. This set of displacements can represent any problem desired and it should be noted that the field is actually needed only for the nodes associated with the distorted ring of elements. In view of this, the asymptotic crack tip displacements can be used everywhere as a suitable field. That is

$$\Delta u_i' = \frac{\Delta K_1^*}{2G_j} \sqrt{\frac{r}{2\pi}} \frac{e^{\kappa}}{(1 + e^{2\pi\epsilon})} f_i (r, \theta, \epsilon, \mu_j), \tag{9}$$
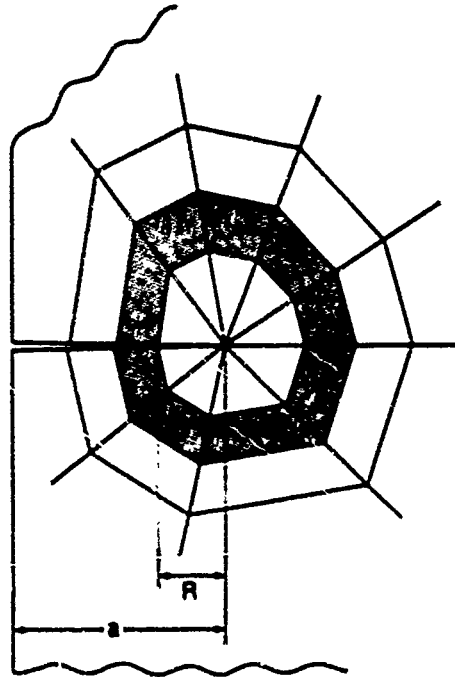
Fig. 2. A typical ring of elements to be distorted in the J-calculations.

where j denotes material 1 or 2 and f_i is described in the Appendix. The vector $\{\Delta u_a\}_I$ is obtained by evaluating (9) at the required nodes. With the value of $\partial\{S\}/\partial a$ already computed for the first computation of J, the calculation in (8) is repeated with the vector $\{u_a\} + \{\Delta u_a\}_I$ used instead of $\{u_a\}$ alone. The result of this calculation, $J + \Delta_I J$, can be shown from (3), (5) and (7) to be such that

$$\Delta_I J = \frac{1}{H}(\Delta K_I^{\bullet 2} + 2K_I^{\bullet}\Delta K_I^{\bullet}). \tag{10}$$

With this rearranged, it follows that

$$K_I^{\bullet} = \frac{H}{2}\frac{\Delta_I J}{\Delta K_I^{\bullet}} - \frac{1}{2}\Delta K_I^{\bullet} \tag{11}$$

and the 2nd term can be neglected compared to the 1st if $\Delta K_I^{\bullet 2}$ is small compared to $H\Delta_I J$.

The procedure can then be repeated for an added vector $\{\Delta u_a\}_{II}$ such that $K_I^{\bullet} = 0$ and $K_{II}^{\bullet} = \Delta K_{II}^{\bullet}$ if desired. These displacements are such that

$$\Delta u_{II} = \frac{\Delta K_{II}^{\bullet}}{2G_j}\sqrt{\frac{r}{2\pi}}\frac{e^{\pi}}{(1 + e^{2\pi})}f_{II}(r, \theta, \varepsilon, \mu_j), \tag{12}$$

where f_{11} is described in the Appendix. Then it follows that the new value from (8) is $J + \Delta_{11}J$ and

$$K_{11}^\circ = \frac{H}{2}\frac{\Delta_{11}J}{\Delta K_{11}^\circ} - \frac{1}{2}\Delta K_{11}^\circ. \tag{13}$$

As an alternative, K_{11}° can be computed directly from (3a) and (3b) given a prior calculation of J and K_1°.

3. Implementation

The finite element calculations for each crack problem were carried out with the ABAQUS program developed by Hibbitt, Karlsson and Sorensen [18]. This code was used once to give the numerical solution of the elastic boundary value problem for which the stress intensity factors are required. The displacement solution, nodal coordinates and element connectivity matrix produced by ABAQUS were written in a sequential output file. A separate program, called KIKII, was developed to perform the virtual crack extension calculations [8] for obtaining K_1° and K_{11}° as described in the previous section. KIKII selects the ring of elements to be distorted and obtains the necessary data from the ABAQUS output file. In the J-calculation subroutine of KIKII, as well as in ABAQUS, eight noded reduced integration isoparametric elements were used. Some calculations were carried out using a DEC Micro-vax II in double precision and others with a CONVEX CI-XP2 in standard precision. In all cases, the computer time to obtain K_1° and K_{11}° from the virtual crack extension calculations is negligible compared to the time needed for obtaining the finite element solution to the boundary value problem for the cracked body.

4. Numerical results

(i) *Examples.* The method was tested on two different examples. The first one consisted of an idealized problem of a circular plane strain region with a straight crack extending along the bimaterial interface from the perimeter to the center of the circle. The outer radius of the region was 10 length units. Poisson's ratio was taken to be 0.3 for the two materials and the ratio of Young's modulus in material 1 to that in material 2 was chosen to be 10. Displacement boundary conditions were imposed around the perimeter. These were exactly compatible with the asymptotic singular crack tip strains for specified values of K_1° and K_{11}°. They were therefore given by the sum of (9) and (12) with ΔK_1° and ΔK_{11}° replaced by the specified K_1° and K_{11}° respectively. In this way it was possible to invoke the exact boundary conditions corresponding to chosen values for K_1°, K_{11}° and Ψ°, where Ψ° is the phase angle of the stress intensity factor (arctan K_{11}°/K_1°).

Several calculations were done using different meshes and rings of elements to test the convergence and mesh dependence. Table 1 shows the results for K_1°, K_{11}° and Ψ° and the percentage error with respect to the exact values of $K_1^\circ = 1.716$, $K_{11}^\circ = 0.208$ and $\Psi^\circ = 6.911$ degrees. It should be noted that the dimensions of K_1° and K_{11}° are stress units times length units raised to power $\frac{1}{2} - i\varepsilon$ as discussed by Rice [17]. Throughout this paper, units are omitted

Table 1 Stress intensities and percentage error for the crack in the circular plane strain region. The reported phase angle Ψ^* is in degrees

Number of elements	KIK/I results			Percentage error		
	K_1^*	K_2^*	Ψ^*	K_1^*	K_2^*	Ψ^*
64	1.7162	0.2079	6.906	0.0120	0.0230	0.0003
512	1.7160	0.2080	6.9110	0.0007	0.0012	0.0
2048	1.7160	0.2080	6.9110	0.0004	0.0020	0.0

implying that an appropriate normalization has been carried out. It is observed that when refined meshes were used, the method gave highly accurate results equal to the exact ones to the fourth decimal digit. The percentage error shown in Table 1 for the very refined meshes represents presumably errors introduced at the stage of the finite element calculation. Even the relatively coarse mesh gives good results.

The second example was the case of a crack of length 2 lying along the interface of a plane stress plate subject to a normal unit tensile stress at the boundaries. One of the meshes used in this analysis to obtain the finite element elasticity solution in a plate of 20 × 20 length units is shown in Fig. 3. It has 160 elements with 64 distributed in the square focused mesh of dimension 2.0 surrounding the crack tip. Symmetry along the line $x = -1.0$ (where the crack tip is at $x = 0$) was used in order to limit the model. Table 2 lists the material properties ratios for three cases solved with this mesh. A comparison between our calculations and the analytical solution for an infinite body derived by Rice and Sih [2] is shown in Table 3. It is observed that this size of plate, 20 × 20, gives an average error of 1 percent
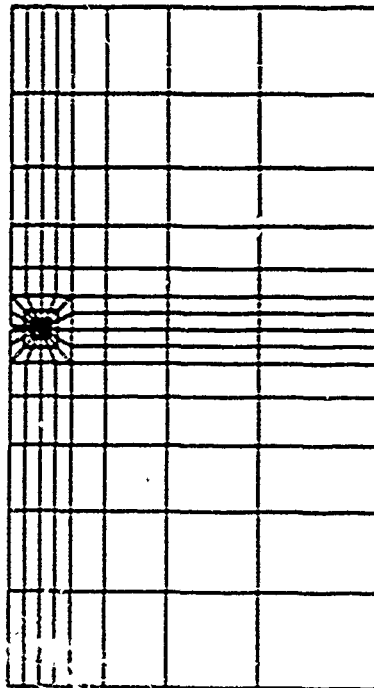
Fig. 3 Finite element mesh used in the second example.

Table 2 Material property ratios for the 20 × 20 length unit plane stress plate

Case no.	$E_1 E_2$	$v_1 v_2$
1	10.	1.
2	22.	0.857
3	100.	1

Table 3 Stress intensities and percentage error for the 20 × 20 lengths units plane stress plate. The reported phase angle Ψ° is in degrees

	Case 1			Case 2			Case 3		
	Ref. [2]	KIKII Progr	% error	Ref. [2]	KIKII progr	% error	Ref. [2]	KIKII progr	% error
K_1°	1.790	1.811	1.15	1.792	1.811	1.09	1.799	1.817	1.01
K_{11}°	-0.217	-0.219	1.10	-0.224	-0.226	0.95	-0.262	-0.264	0.89
Ψ°	-6.897	-6.894	0.05	-7.134	-7.124	0.13	-8.304	-8.291	0.13

Table 4 Results and percentage error for the 40 × 40 length units plate

	K_1°	K_2°	Ψ°
KIKII progr	1.7953	-0.2172	-6.8995
% error	0.27	0.30	0.05

Table 5 Percentage error for different rings for the 40 × 40 length units plate

	Percentage error		
R.a	K_1°	K_2°	Ψ°
0.04	0.26	0.36	0.09
0.09	0.27	0.32	0.05
0.16	0.27	0.32	0.04
0.34	0.28	0.32	0.04
0.44	0.28	0.29	0.02

Table 6 Percentage error for different values of ΔK_1° and ΔK_{11}°

$\dfrac{\Delta K^\circ}{\sqrt{J/H}}$	Circular plate		40 × 40 plate	
	Percentage error		Percentage errors	
	K_1°	K_{11}°	K_1°	K_{11}°
10^{-8}	0.730×10^{-3}	0.127×10^{-3}	0.286	0.329
10^{-6}	0.730×10^{-3}	0.127×10^{-3}	0.286	0.329
10^{-2}	0.729×10^{-3}	0.128×10^{-3}	0.286	0.329
10^{-1}	0.722×10^{-3}	0.135×10^{-3}	0.286	0.328
10^1	0.251×10^{-4}	0.920×10^{-3}	0.279	0.217
10^3	0.748×10^{-1}	0.794	0.351	10.9

for the stress intensity factors. The same crack length in a larger plate of 40 × 40 length units and using the ratio of material properties of Case 1, gives the results shown in Table 4. The mesh for this problem contained 426 elements. The average error in K_1° and K_{11}° is now 0.3 percent. Based on these results, one can conclude that the finite size approximation for the infinite plate makes a large contribution to the overall error shown in Table 3 and relatively little of the error is due to the technique used to evaluate the stress intensity factor. Again, good results are obtained with a coarse mesh.
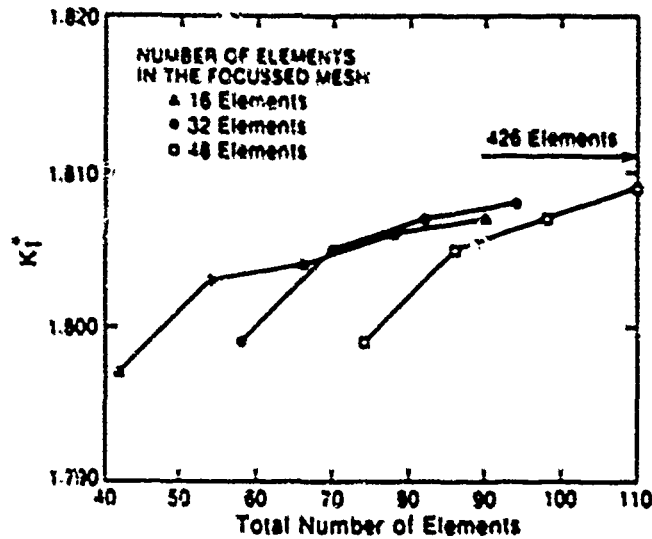
Fig. 4 Effect of the number of elements on the calculated value of K_I^*
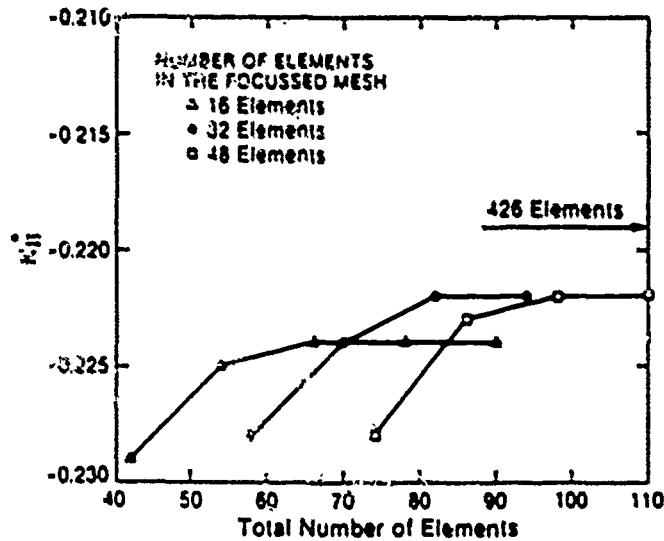
Fig. 5. Effect of the number of elements on the calculated value of K_{II}^*

The effect of choosing different rings of elements for distortion in the virtual crack extension on the accuracy of K_I^*, K_{II}^* and Ψ^* was investigated for the 40 × 40 plate discretized in 426 elements. The percentage error compared to the infinite plate solution [2] for five different cases is shown in Table 5, in terms of R/a, the distance from the crack tip to the distorted element ring measured along the x-axis in Fig. 2 normalized by the crack length a. The accuracy of the method is independent of the choice of the ring to be distorted. Therefore, any ring can be taken, except the one containing the crack tip nodes. In this case

a singularity is introduced by the displacement increments calculated by (9) and (14) making it unsuitable for the procedure.

Convergence as a function of the stress intensity factor increments was evaluated by examining the trend in percentage error of K_I^* and K_{II}^* for both examples investigated. Table 6 shows that the method is accurate for increment magnitudes smaller than 10 $\cdot \sqrt{J\bar{h}}$, for both examples presented. Even with large changes ΔK^* the results are good as long as (11) is accurately implemented.

The effect of discretization on convergence was investigated for the 20 × 20 plate using the material properties of Case 1 and mesh arrangement similar to Fig. 3. The number of elements focused at the crack tip was kept constant for a given test and the total number of elements increased by equal increments. A ring of elements within the focused mesh was distorted for the J-calculations. The results were compared with $K_I^* = 1.811$ and $K_{II}^* = -0.219$, obtained for a dense mesh of 426 elements. As Figs. 4 and 5 show, K_I^* and K_{II}^* became insensitive to the number of focussed elements as the total number of elements became large. Conversely, a lower accuracy was obtained with a small number of focussed elements. It suggests that the discretization in the near tip region has an important role in the accuracy of the method and this must be balanced with the total number of elements required for accurate results.

5. Crack surface displacement method

In this section, a different approach is used to evaluate the stress intensity factor. This method, unlike the one presented earlier, is semi-energetic; i.e. J is computed in terms of the rate of change of potential energy, whereas the phase angle Ψ^* is obtained through the crack surface displacements. The finite element node at which Ψ^* is evaluated is chosen through an internal consistency check which balances the numerical error on the displacements in the near tip region. Results obtained by this method with fine finite element meshes were in excellent agreement with those obtained via the other method. An outline of the crack surface displacement (CSD) method and its comparison with the other method are presented below.

As discussed earlier in this paper, the small strain elastic solution for the stresses in the region of an interfacial crack tip has the oscillatory character shown by (1). In accordance with (1), the crack surface displacements exhibit a similar oscillatory solution, i.e. for plane strain

$$\Delta u_2 + i\Delta u_1 = \frac{2[(1 - \nu_1)/G_1 + (1 - \nu_2)/G_2]}{(1 + i2\epsilon)\cosh(\pi\epsilon)}(K_I^* + iK_{II}^*)\sqrt{\frac{r}{2\pi}}r^{i\epsilon},\tag{14}$$

where r is the distance from the crack tip. The symbol Δ on the left hand side of (14) now denotes the relative displacements of two initially neighboring points on the crack surfaces behind the tip as shown by Fig. 6. For the sake of the analysis we can rewrite the above equation as follows,

$$|\Delta u| e^{i\phi} = \alpha h^{1/2}(K_I^* + iK_{II}^*) h^{i\epsilon} \left(\frac{r}{h}\right)^{1/2-i\epsilon} e^{-i\psi},\tag{15}$$

Fig 6. The near tip crack surface displacements used in CSD method. In this figure, $\Delta x = \Delta u_x = \Delta u_y + i\Delta u$. used in (14) and (19).

where

$$x = \frac{2[(1 - \nu_1)/G_1 + (1 - \nu_2)/G_2]}{\sqrt{2\pi(1 + 4\epsilon^2)}\cosh(\pi\epsilon)}.$$ (16)

$|\Delta u|$ and ϕ are the modulus and argument respectively of the complex quantity $\Delta u_2 + i\Delta u_1$, and β is the principal argument of the complex number $1 + i2\epsilon$. As shown by (15) the product $(K_I^* + iK_{II}^*)h^{i\epsilon}$, h being a characteristic specimen dimension, constitutes a new rotated quantity $\hat{K}_I^* + i\hat{K}_{II}^*$, with the same modulus $|K|$ as $K_I^* + iK_{II}^*$ but an augmented phase angle Ψ^* such that $\Psi^* = \Psi^* + \epsilon \ln h$. The new phase angle Ψ^*, unlike Ψ^*, is in fact insensitive to the choice of length unit. This becomes clearer when we rewrite (15) with all complex quantities expressed in polar form, i.e.

$$|\Delta u| e^{i\phi} = xh^{1/2}\left(\frac{r}{h}\right)^{1/2}[(K_I^*)^2 + (K_{II}^*)^2]^{1/2} e^{i\left(\Psi^* + \epsilon \ln \frac{r}{h} - \beta\right)}.$$ (17)

which yields the following expressions for the scale insensitive phase angle Ψ^* and the modulus

$$\Psi^* = \phi - \epsilon \ln \frac{r}{h} + \beta$$ (18)

and

$$|K| = [(K_I^*)^2 + (K_{II}^*)^2]^{1/2} = \frac{[2\pi(1 + 4\epsilon^2)(\Delta u_1^2 + \Delta u_2^2)]^{1/2} \cosh(\pi\epsilon)}{2[(1 - \nu_1)/G_1 + (1 - \nu_2)/G_2]h^{1/2}(r/h)^{1/2}}.$$ (19)

For example, h could be the width of the specimen, rendering r/h to be a distance from the crack tip relative to the width of the specimen. The utility of this approach will become clear below.

Fig. 7. The trends in the modulus |K| computed from the nodal displacements (dash line) and the energy release rate (solid line). In this plot, / is the crack length, $\Delta u_1 = \Delta u_1$ and $\Delta u_2 = \Delta u_2$.

It is clear that with use of (18) and (19) finite element results for displacements (Δu_1, Δu_2) along the crack surface can be used to obtain an estimate of both the phase angle Ψ^* and the modulus |K|. Calculations of this type were carried out for the center cracked plate examined earlier in this work. It was found that the |K| estimate, shown by the dash line in Fig. 7, varies almost linearly with r/h in the region close to the crack tip where r is now the position at which the displacements were used for (18) and (19) and h is a characteristic length taken to be one length unit. On the other hand, a very accurate estimate of |K| can be obtained from the J-integral calculated by the virtual crack extension method. The relationship given by (3a) and (3b), is

$$|K| = \left[\frac{J}{2[(1 - v_1)/G_1 + (1 - v_2)/G_2]} \right]^{1/2} 2 \cosh (\pi\varepsilon). \tag{20}$$

The virtual crack extension result for J and consequently for |K| has been found to be independent of r/h where r is the radial position of the ring of elements distorted for this calculation. The results are shown in Fig. 7 with a solid line. As can be seen, the two estimates of |K| coincide at a distance r_s from the crack tip. Therefore, in the neighborhood r_s, nodal displacements give a good estimate of |K|. It is likely that the corresponding estimate for Ψ^* given by (18) is also good at that location. Indeed, for the case of the center cracked plate, the predicted values of |K| and Ψ^* were in remarkable agreement with those found using the energy method (difference <0.5% for K_1^*) as shown in Table 7. Thus the CSD method is henceforth implemented by calculating J by the virtual crack extension method and finding

Table 7 Results for the 20 × 20 and 40 × 40 length units center cracked plate obtained via the CSD and the energy method respectively. The % error column denotes the deviation of the CSD method results from those of the energy results.

Length units		CSD method	Energy method	% error
20	K_I^*	1.810	1.811	0.05
	K_{II}^*	−0.224	−0.219	2.28
	Ψ^* (rad)	−0.123	−0.120	2.21
40	K_I^*	1.795	1.7953	0.017
	K_{II}^*	−0.222	−0.2172	2.21
	Ψ^* (rad)	−0.123	0.1204	2.20

the position on the crack for which (19) agrees with (20), where J in (20) is the virtual crack extension value. At that location, the crack surface displacements are used in (18) to compute the phase angle. The factors K_I^* and K_{II}^* are then computed by finding the real and imaginary parts of $|A'|\,e^{+\star}$.

6. Comparison of the methods

The two methods were used to extract K_I^* and K_{II}^* from finite element results for the prenotched bimaterial beam in flexure shown in Fig. 8. Due to geometric and elastic symmetry with respect to its midsection, the analysis was carried out in the right half of the beam. The finite element mesh used is shown in Fig. 9 and as can be seen, in the crack tip region the mesh was focused with eight 16-element rings. The total number of elements was 1120. These elements were 8-noded isoparametric with 4-stations for the integration of the element stiffness. A series of results were obtained and tabulated (see Tables 8, 9, 10) for various crack lengths. The height h of the beam (see Fig. 8) was chosen as the characteristic specimen dimension and the values reported in the above tables correspond to the rotated stress intensity factor $K_I^* + iK_{II}^* = (K_I^* + iK_{II}^*)/r^*$. In addition in the same tables, the stress intensity components were non-dimensionalized by $Pl/bh^{3/2}$ and the energy release rate by $P^2l^2(1 - v_2^2)/b^2h^3E_2$ where P is the applied load, l is the spacing between the inner and outer loading points, b is the width of the beam and E_1 the elastic modulus of the top layer (see Fig. 8).
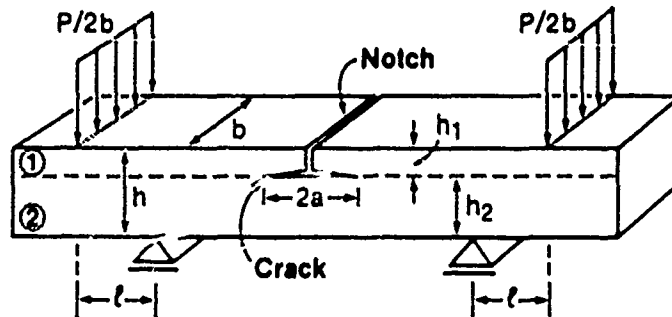
Fig. 8. The four-point bending beam with two symmetrical interfacial cracks used in the comparison of the CSD and energy methods.

a) The Deformed Body $E_2/E_1 = 100$
$\nu_2 = \nu_1 = 0.3$

b) The Deformed Crack Tip Region
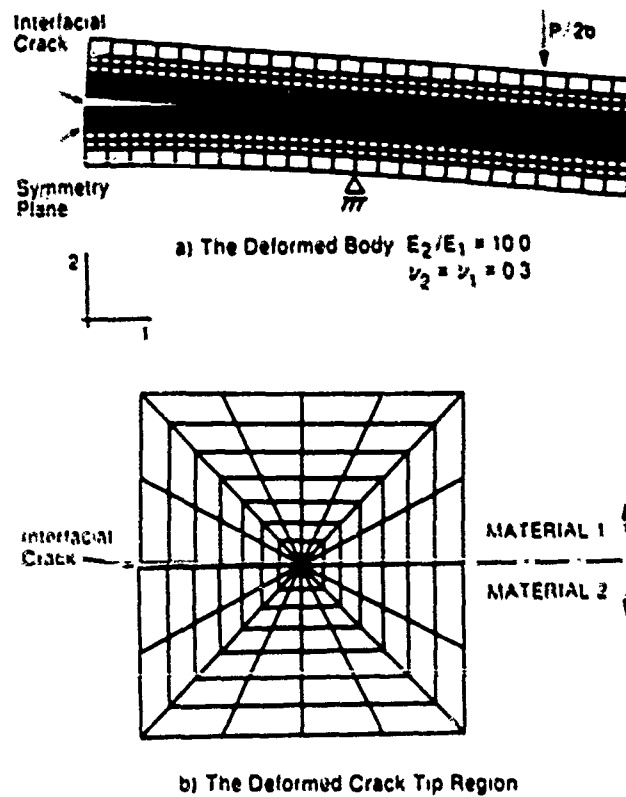
Fig. 9. The mesh used in the finite element calculations.

Table 8. The trends of the energy release rate J and normalized complex stress intensity factor $K^\circ h^\epsilon$ with respect to the interfacial crack length for the homogeneous case, i.e. $E_2/E_1 = 1.0$ and $\nu_2 = \nu_1 = 0.3$. Columns marked by 1 and 2 correspond to results obtained by the energy and CSD methods respectively.

E_2/E_1	a/l	J	$\hat{K}_I^\circ$	$\hat{K}_I^\circ$	$\hat{K}_{II}^\circ$	$\hat{K}_{II}^\circ$	Ψ° (degrees)	Ψ° (degrees)
			1	2	1	2	1	2
1.0	0.0938	10.2097	2.4501	2.4645	2.0513	2.0337	39.9363	39.5297
1.0	0.1563	10.4302	2.4478	2.4610	2.1074	2.0914	40.7262	40.3582
1.0	0.3125	10.4814	2.4484	2.4604	2.1201	2.1042	40.8904	40.5383
1.0	0.4688	10.4780	2.4487	2.4601	2.1200	2.1038	40.8849	40.5352
1.0	0.6250	10.4707	2.4498	2.4597	2.1196	2.1025	40.8674	40.5233
1.0	0.7813	10.4766	2.4535	2.4619	2.1194	2.1013	40.8206	40.4810
1.0	0.9375	10.5233	2.4626	2.4710	2.1194	2.1018	40.7160	40.3834
1.0	1.0938	10.5751	2.4736	2.4810	2.1208	2.1023	40.6090	40.2765
1.0	1.2500	10.4256	2.4477	2.4516	2.1234	2.1012	40.9415	40.5992
1.0	1.4063	9.3495	2.2530	2.2569	2.0870	2.0630	42.8090	42.4303
1.0	1.5625	6.7700	1.8376	1.8359	1.8688	1.8437	45.4817	45.1212

Table 9 The trends of the energy release rate *J* and normalized complex stress intensity factor K^*k^* with respect to the interfacial crack length for elastic moduli ratios $E_2/E_1 = 10.0$ and Poisson's ratios $v_2 = v_1 = 0.3$. Columns marked by 1 and 2 correspond to results obtained by the energy and CSD methods respectively

E_2/E_1	a/l	J	K_I^* 1	K_I^* 2	K_{II}^* 1	K_{II}^* 2	ψ^* (degrees) 1	ψ^* (degrees) 2
10.0	0.0938	6.4022	0.7830	0.7887	0.7866	0.7805	45.1305	44.6989
10.0	0.1563	6.4681	0.7701	0.7751	0.8071	0.8020	46.3443	45.9903
10.0	0.3125	6.5033	0.7685	0.7711	0.8131	0.8082	46.6133	46.2719
10.0	0.4688	6.4912	0.7678	0.7722	0.8126	0.8076	46.6239	46.2829
10.0	0.6250	6.4626	0.7662	0.7702	0.8115	0.8061	46.6427	46.3041
10.0	0.7813	6.4088	0.7628	0.7664	0.8090	0.8033	46.6860	46.3491
10.0	0.9375	6.3065	0.7554	0.7590	0.8039	0.7983	46.7805	46.4460
10.0	1.0938	6.0824	0.7388	0.7421	0.7928	0.7869	47.0178	46.6806
10.0	1.2500	5.5802	0.6997	0.7018	0.7690	0.7621	47.7024	47.3574
10.0	1.4063	4.5560	0.6114	0.6129	0.7150	0.7076	49.4657	49.1005
10.0	1.5625	3.0571	0.4800	0.4795	0.6066	0.5985	51.6449	51.3016

Table 10 The absolute percentage difference between CSD and energy method predictions. The energy results were used as datum.

a/l	$E_2/E_1 = 1.0$ $v_2 = v_1 = 0.3$			$E_2/E_1 = 10.0$ $v_2 = v_1 = 0.3$		
	K_I^* % error	K_{II}^* % error	ψ^* % error	K_I^* % error	K_{II}^* % error	ψ^* % error
0.0938	0.5877	0.8580	1.0181	0.7280	0.7755	0.9563
0.1563	0.5393	0.7592	0.9036	0.6493	0.6319	0.7854
0.3125	0.4901	0.7500	0.8611	0.5986	0.6026	0.7367
0.4688	0.4656	0.7642	0.8553	0.5731	0.6153	0.7333
0.6250	0.4041	0.8068	0.8420	0.5221	0.6654	0.7259
0.7813	0.3424	0.8540	0.8319	0.4719	0.7046	0.7216
0.9375	0.3411	0.8304	0.8169	0.4766	0.6966	0.7150
1.0938	0.2992	0.8723	0.8188	0.4467	0.7442	0.7172
1.2500	0.1593	1.0455	0.8361	0.3001	0.8973	0.7232
1.4063	0.1731	1.1500	0.8846	0.2453	1.0350	0.7383
1.5625	0.0925	1.3431	0.7926	0.1042	1.3352	0.6647

A series of results were obtained for non-dimensional interfacial crack lengths in the interval $0.0938 \le a/l \le 1.5625$, the inner loading point being at $a/l = 1.5$. The results for the homogeneous case, i.e. $E_2/E_1 = 1.0$ are shown in Table 8 and those for $E_2/E_1 = 10.0$ in Table 9. The Poisson's ratio for all materials in both cases was chosen to be $v_1 = v_2 = 0.3$. The column indices 1 and 2 correspond to the results obtained via the energy method and CSD method respectively. In both cases, J was obtained from the stiffness derivative method [8]. For cracks in the pure bending region between the inner supports, the J values were found to be in excellent agreement with estimates obtained by Charalambides, Lund, Evans and McMeeking [19] from energy release considerations and beam bending theory.

As shown in Table 8, the K_I^* values obtained via the CSD method were consistently higher than those predicted by the energy method, and as a result, in order to maintain the same $|K|$ the corresponding K_{II}^* values predicted by the CSD method were consistently lower than those of the energy method. The absolute percentage difference denoted by % error for both K_I^* and K_{II}^* is shown in Table 10. In addition, in the same table we report on the difference in ψ^* between the two methods. This difference was found to be of the order of 1 percent

for short cracks and 0.7 percent for cracks in the region between the inner and outer load points where the bending moment varies with crack length. In the case of a homogeneous beam, this region is dominated by the shear component stress intensity factor K_{II}. As shown in Tables 9 and 10, for both the homogeneous and the bimaterial cases, the deviation in $\hat{K}_{II}^{\circ}$ values obtained via the above methods increases as the crack approaches the transient moment region and is of the order of 1 percent for cracks between the inner and outer loading lines. On the other hand the deviation in $\hat{K}_{I}^{\circ}$ decreases as the crack length increases and is of the order of 0.1 percent for cracks between the inner and outer loading lines.

7. Discussion

The internal consistency check in the CSD method of using the location where the displacement based values of J agrees with the virtual crack extension J value does bring the results obtained very close to the predictions entirely based on the virtual crack extension method. Although this brings confidence to the results of both methods, this comparison is pertinent only to focused meshes with a reasonably well refined crack tip region.
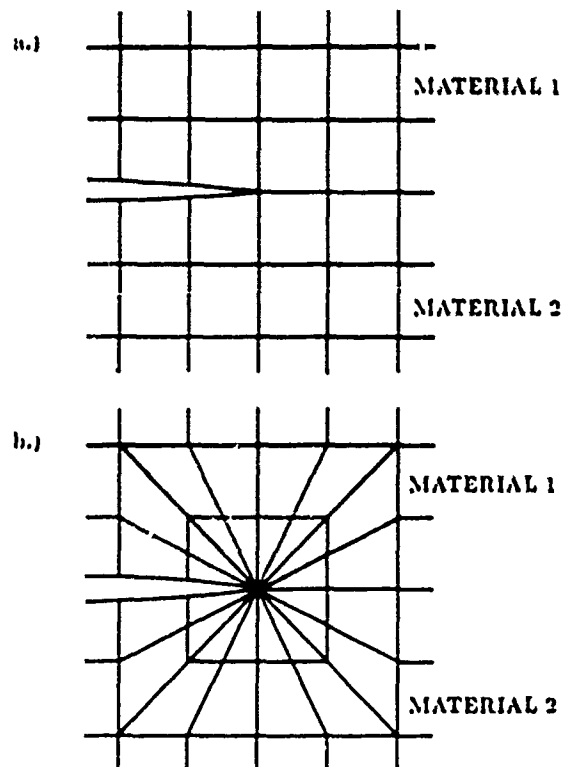
CRACK TIP REGIONS

Fig. 10. The rectangular (10a) and the focused (10b) crack tip mesh used to examine the crack tip mesh type sensitivity of the CSD and energy methods.

Table 11 Selective results for two different crack lengths obtained by the CSD method. Columns 1 and 2 correspond to the crack tip mesh type shown in Fig. (10b) and Fig. (10a) respectively.

E_2/E_1	all	J 1	J 2	K_I^* 1	K_I^* 2	K_{II}^* 1	K_{II}^* 2	ψ^* (degrees) 1	ψ^* (degrees) 2
1.0	0.9375	10.5233	10.5692	2.4710	2.5060	2.1018	2.0710	40.3834	39.5513
1.0	1.4063	9.3495	9.4494	2.2569	2.3990	2.0630	2.0330	42.4303	41.4879
10.0	0.9375	6.3068	6.2583	0.7590	0.8370	0.7983	0.6840	46.4460	39.2647
10.0	1.4063	4.5560	4.4524	0.6129	0.6690	0.7076	0.6130	49.1005	42.7254

Table 12 The absolute percentage difference marked as % error between the CSD results for the mesh type 10a and 10b, the second being the datum.

E_2/E_1	all	J % error	K_I^* % error	K_{II}^* % error	ψ^* % error
1.0	0.9375	0.4362	1.4974	1.4654	2.0608
1.0	1.4063	1.0685	1.8654	1.4542	2.2210
10.0	0.9375	0.8005	10.2767	14.3179	15.4616
10.0	1.4063	2.2743	9.1532	13.3691	12.9836

For example, Charalambides et al. [20] reported K^* values for the same specimen geometry using a rectangular mesh in conjunction with the CSD method. In that analysis the elements used were also 8-noded isoparametric and the crack tip region was discretized as shown schematically in Fig. 10a, whereas the results reported earlier in this paper were obtained for the focused crack tip mesh shown schematically in Fig. 10b. Overall, the mesh containing the near tip region shown in Fig. 10a had a total of 1232 elements versus a total of 1120 in the focused mesh. Selective results for two crack lengths and two moduli ratios, i.e. E_2/E_1 = 1.0, 10.0 are presented in Tables 11 and 12. The columns marked by the index 1 correspond to the focused mesh results and those marked by 2 represent the rectangular mesh results. Here, as before, we use the same non-dimensional quantities. In Table 12, the % error mark denotes the absolute percentage difference of the rectangular mesh results relative to those of the focused mesh. As we can see in the table, the rectangular mesh predictions are most in error compared to the results of the energy method although bounded by a maximum deviation of 15 percent. It is also of some interest to notice that the J estimates from the rectangular mesh are slightly different from those obtained using a focused mesh. This may be an indication of the expected mesh sensitivity of the energy method when used in conjunction with a rectangular mesh. In that respect these results are very encouraging as far as the energy method is concerned since the J estimates appear to be rather mesh insensitive.

A previous analysis (see Figs. 4 and 5) showed that the energy method, unlike the CSD method, depends very little on the number of finite elements or the element ring used, especially when the total number of elements is greater than 50. These results, in light of the rectangular mesh results, suggest that the energy method is indeed very powerful in estimating stress intensities at interfacial crack tips even when used in conjunction with coarse meshes. On the other hand, the CSD method when used with focused meshes, also provides a reliable technique for extracting the stress intensity factors. However, in the latter case, one should be aware of the dependency of the results on the number of elements in the finite element mesh and that it is less accurate for meshes with less than 300 8-noded isoparametric elements.

252   *P.P.L. Matos et al.*

8. Conclusions

Two methods for evaluating the stress intensity factors for crack tips were presented. An advantage of the methods is that they are general and can be used for cracks on bimaterial interfaces. A finite element solution for the crack problem is required in both methods. Thereafter, the computations for the evaluation of the stress intensity factors consist only of a very small fraction of those required to obtain the finite element solution for the crack problem. As a result predictions of high accuracy can be obtained at minimum cost. The crack surface displacement method was found to be more sensitive to mesh refinement and mesh-type in the tip region than the energy method. This is an expected result since the J-calculations are, in general, mesh insensitive. This makes the energy method a more attractive one especially when the crack problem finite element solution is obtained using a coarse mesh.

Appendix

The components of the functions $f_{\rm I}(r, \theta, \varepsilon, \mu_j)$ and $f_{\rm II}(r, \theta, \varepsilon, \mu_j)$ that introduce the oscillatory characteristic of the near crack tip displacement field are obtained as follows:

$$f_{1\rm I} = D_j + 2\delta_j \sin\theta \sin\Psi \tag{A1}$$

$$f_{2\rm I} = -C_j - 2\delta_j \sin\theta \cos\Psi \tag{A2}$$

and

$$f_{1\rm II} = -C_j + 2\delta_j \sin\theta \cos\Psi \tag{A3}$$

$$f_{2\rm II} = -D_j + 2\delta_j \sin\theta \sin\Psi. \tag{A4}$$

D_j, C_j, d_j and Ψ are defined as

$$\delta_1 = e^{-(\pi-\theta)\varepsilon} \quad \delta_2 = e^{(\pi-\theta)\varepsilon}$$

$$\Psi = \varepsilon \log r + \frac{\theta}{2}$$

$$D_i = \beta \gamma_i \cos \frac{\theta}{2} + \beta' \gamma_i' \sin \frac{\theta}{2}$$

$$C_i = \beta' \gamma_i \cos \frac{\theta}{2} - \beta \gamma_i' \sin \frac{\theta}{2}$$

$$\beta = \frac{0.5 \cos (\varepsilon \log r) + \varepsilon \sin (\varepsilon \log r)}{0.25 + \varepsilon^2}$$

$$\beta' = \frac{0.5 \sin (\varepsilon \log r) - \varepsilon \cos (\varepsilon \log r)}{0.25 + \varepsilon^2}$$

$$\gamma_i = \mu_i \delta_i - \frac{1}{\delta_i} \quad \gamma_i' = \mu_i \delta_i + \frac{1}{\delta_i},$$

where (r, θ) are the polar coordinates as shown in Fig. 1, ε is the bimaterial constant as defined before and j is the material index.

References

1. M.F. Kanninen and C.H. Popelar, *Advanced Fracture Mechanics*, Oxford University Press, New York (1985).
2. J.R. Rice and G.C. Sih, *Journal of Applied Mechanics* 32 (1965) 418–423.
3. F. Erdogan, *Journal of Applied Mechanics* 30 (1963) 232–236, and *ibid* 32 (1965) 403–410.
4. A.H. England, *Journal of Applied Mechanics* 32 (1965) 400–402.
5. J.R. Rice, *Journal of Applied Mechanics* 35 (1968) 379–386.
6. J.R. Willis, *Journal of the Mechanics and Physics of Solids* 19 (1971) 353–368.
7. R.H. Gallagher, in *Numerical Methods in Fracture Mechanics*, Proceedings of the First International Conference, Swansea (1978) 1–25.
8. D.M. Parks, *International Journal of Fracture* 10 (1974) 487–502.
9. H. Ishikawa, H. Kitigawa and H. Okamura, in *Mechanical Behavior of Materials*, Proceedings of the Third International Conference held in Cambridge, England, K.J. Miller and R.F. Smith (eds.), Vol. 3, Pergamon (1979) 447–455.
10. H.D. Bui, *Journal of Mechanics and Physics of Solids* 31 (1983) 439–448.
11. J.F. Yau, S.S. Wang and H.T. Corten, *Journal of Applied Mechanics* 47 (1980) 335–341.
12. M. Stern, E.B. Becker, and R.S. Dunham, *International Journal of Fracture* 12 (1976) 359–368.
13. F.H.K. Chen and R.T. Shield, *Zeitschrift fuer angewandte Mathematik und Physik* 28 (1977) 1–22.
14. S.S. Wang and J.F. Yau, *AIAA Journal* 19 (1981) 1350–1356.
15. J.R. Rice, private communication, 1986.
16. C.F. Shih and R.J. Asaro, "Elastic-Plastic Analysis of Cracks on Bimaterial Interfaces Part I: Small Scale Yielding", Report (1987).
17. J.R. Rice, *Journal of Applied Mechanics* 55 (1988) 98–103.
18. H.D. Hibbitt, B. Karlsson and E.P. Sorensen, *User's Manual*, Hibbitt, Karlsson and Sorensen Inc., Providence RI (1984).
19. P.G. Charalambides, J. Lund, A.G. Evans and R.M. McMeeking, *Journal of Applied Mechanics* 56 (1989) 77–82.
20. P.G. Charalambides, J. Lund, A.G. Evans and R.M. McMeeking, "A Test Specimen for Determining the Fracture Resistance of Bimaterial Interfaces," *The Processing and Mechanical Properties of High Temperature/High Performance Composites*, Annual Report, Department of Materials, University of California at Santa Barbara (1987).

Résumé. On présente une méthode numérique en vue d'obtenir les valeurs de K_I^*, K_{II}^* et K_{III}^* relatives à la solution élastique d'application à l'extrémité d'une fissure d'interface sujette à un état de contraintes général. La méthode repose sur l'évaluation de l'intégrale J par la technique d'extension virtuelle de la fissure. On peut ensuite obtenir les intensités de contraintes individuelles à partir de calculs de J subséquents, correspondant à des perturbations introduites par de petits accroissements des facteurs d'intensité de contraintes.

Les calculs sont accomplis par la méthode des éléments finis, mais, par rapport aux calculs à mettre en oeuvre dans le problème des valeurs aux limites, il ne faut procéder qu'à quelques calculs supplémentaires.

On présente des résultats très précis pour le cas d'une fissure dans un interface entre deux matériaux, et on les compare avec ceux provenant d'autres méthodes d'évaluation des facteurs d'intensité de contraintes.

En particulier, on fait une comparaison pour des facteurs d'intensité de contraintes obtenus en calculant J par la méthode d'extension virtuelle d'une fissure, mais en séparant les modes selon le rapport des déplacements de la surface de la fissure.

Les deux techniques fonctionnent de manière satisfaisante avec des maillages fins d'éléments finis; cependant, les résultats suggèrent que la méthode qui repose entièrement sur les évaluations de l'intégrale J peut être utilisée afin d'obtenir des résultats fiables dans les réseaux à mailles grossières.

MATERIALS
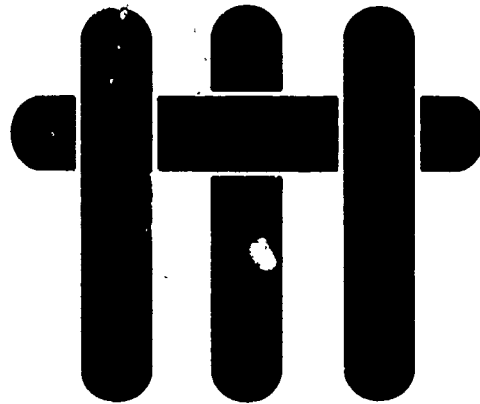
EFFECTS OF NON-PLANARITY ON THE

MIXED MODE FRACTURE RESISTANCE

OF BIMATERIAL INTERFACES

A. G. Evans

Materials Department
College of Engineering
University of California, Santa Barbara
Santa Barbara, California 93106

and

J. W. Hutchinson

Division of Applies Sciences
Harvard University
Cambridge, Massachusetts 02138

EFFECTS OF NON-PLANARITY ON THE MIXED MODE FRACTURE RESISTANCE OF BIMATERIAL INTERFACES

A. G. EVANS[1] and J. W. HUTCHINSON[2]

[1]Materials Department, University of California, Santa Barbara, CA 93106 and [2]Division of Applied Sciences, Harvard University, Cambridge, MA 02138, U.S.A.

Abstract—The effects of non-planarity on the fracture resistance locus of interfaces has been investigated using a simple model of contacting facets along the crack surface. The contacts resist the motion of the crack surface by means of friction and locking and thereby modify the energy release rate G^* at the crack front. The modified G^* governs the effect of the contacting facets on the overall interface fracture resistance, G_i. The trends in G_i with phase angle of loading are found to be influenced largely by a non-dimensional parameter that determines the length of the contact zone. This parameter is, in turn, dependent on the amplitude of the undulations on the fracture interface as well as its intrinsic fracture resistance.

Résumé—Nous avons étudié les effets de la non-planéité sur les lieux de résistance à la rupture des interfaces en utilisant un modèle simple de facettes en contact le long de la surface de la fissure. Les contacts freinent le mouvement de la surface de la fissure par friction et blocage, et ils modifient donc la vitesse de perte d'énergie G^* en tête de fissure. Cette vitesse G^* modifiée régit l'effet des facettes en contact sur la résistance à la rupture interfaciale totale, G_i. Les variations de G_i avec l'angle de phase de la mise en charge sont fortement influencées par un paramètre sans dimension qui détermine la longueur de la zone de contact. Ce paramètre, par contre, dépend de l'amplitude des ondulations sur l'interface de rupture comme de sa résistance intrinsèque à la rupture.

Zusammenfassung—Der Einfluß der Unebenheit auf den Ort des Bruchwiderstandes von Grenzflächen wurde mit einem einfachen Modell kontaktierender Facetten entlang der Bruchfläche untersucht. Die Kontakte widerstehen der Bewegung der Rißoberfläche durch Reibung und Verankerung und verändern daher die Freisetzungsrate der Energie G^* an der Rißfront. Die modifizierte Rate G^* beherrscht den Einfluß der kontaktierenden Facetten auf den gesamten Bruchwiderstand der Grenzfläche G_i. Die Abhängigkeiten von G^* von dem Phasenwinkel der Belastung werden größtenteils durch einen dimensionslosen Parameter beeinflußt, der die Länge der Kontaktzone bestimmt. Dieser Parameter hängt wiederum von der Amplitude der Wellungen auf der Bruch-Grenzfläche und deren intrinsischem Bruchwiderstand ab.

1. INTRODUCTION

Many important interface fracture problems involve mixed mode (shear and opening) displacements along the crack surfaces, as exemplified by thin film decohesion [1-3] and fiber debonding in composites [4]. Subject to such displacements, interface fracture must be influenced by non-planarity of the interface and by the phase angle of loading, $\Psi = \tan^{-1}(K_{II}'/K_I')$, where K_I' and K_{II}' are the Mode I and Mode II stress intensity factors. Typical interfaces are non-planar [5] (Fig. 1) and crack surface contact either at undulations or at facets along the interface crack can have an effect on the overall fracture resistance of the interface, G_i, especially at large phase angles. Such phase angle effects are illustrated in Fig. 2, which indicates the results of a fracture test on Al_2O_3 bonded with Ti. The upper, interface failure was caused by applied loading, with $\Psi \approx 0$. The lower crack in the Al_2O_3 parallel to the interface formed subsequently, caused

by the residual stress in the Ti acting as a thin film. For this case, $\Psi \approx \pi/4$ [6] at the interface crack, whereupon the crack is diverted into the Al_2O_3 rather than propagating along the interface.

Trends in G_i with phase angle of loading are predicted in the present study for the case of a faceted interface. The simplest model of the process that provides insightful preliminary conclusions, illustrated in Fig. 3, consists of kinks along the crack surface. When the crack surfaces contact at the kinks, the stress intensities K^* at the crack front differ from the applied values to an extent governed by the kink angle, the kink amplitude and the friction coefficient. The trends in K^* with these variables provide one contribution to the increase in interface fracture resistance with phase angle of loading, as elaborated in the present article. Other possible influences on G_i, such as crack front deflections, plasticity etc. are not considered in this study.

AM 37 3—K

(a)
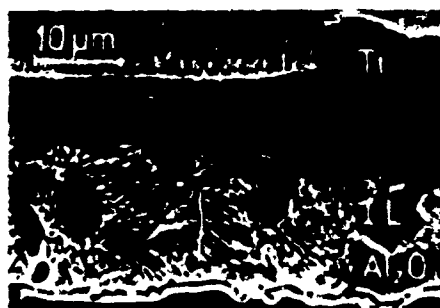
(b)

Fig. 1 Non-planar metal ceramic interfaces (courtesy M. Rühle). (a) Interface facets in the system Nb Al₂O₃ (b) Interface undulations in the system Ti Al₂O₃, with interphases of TiAl and Ti₃Al. Cracking occurs along the Al₂O₃ TiAl interface.
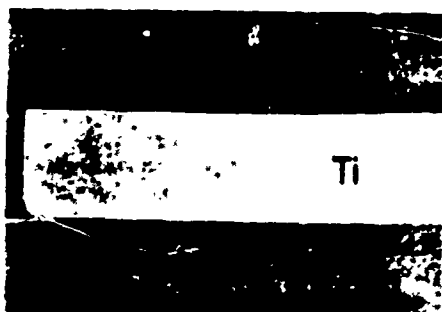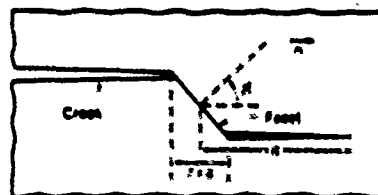

Fig. 2 The results of a fracture test on a Ti Al₂O₃ system. The upper, interface crack formed first upon loading. The lower crack in the Al₂O₃ formed subsequently, because of residual stress in the Ti (courtesy M. Rühle).
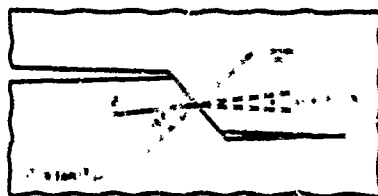
Fig. 3. The crack kink model used to analyse effects of crack surface contact. The loading in this case is $A_2 < 0$ and $A_1 > 0$.

2. THE BASIC MODEL

The model is developed for a thin interface between elastically homogeneous bodies with identical elastic properties. However, the general trends should be applicable to bimaterial interfaces, albeit that complex stress intensities should then be used [7]. The basic geometry involves a single kink (Fig. 3(a)) at angle β, along the crack surfaces at a distance, d, from the crack front. This geometric choice simulates each contact along a multiply facetted interface.

When the phase angle of loading allows contact at the kink, Coulomb friction is assumed to obtain with a friction angle

$$\sigma = \arctan \mu \qquad (1)$$

where μ is the friction coefficient. Otherwise, elasticity exists throughout. An approximation for the effect of contact on the crack tip field involves an inclined force, F, acting on the surfaces of a planar crack (Fig. 3(c)), with the inclination angle ω governed by σ and the kink angle β, (assuming that the faces of the facet are either sliding or on the verge of sliding) where

$$\omega = \beta - \phi \qquad (2)$$

The derivation which follows assumes $0 < \beta < \pi/2$, as in Fig. 3. For kink angles in the range $\pi/2 < \beta < \pi$, all equations continue to be valid if ϕ is replaced by

$-\phi$ corresponding to the switch in direction of the friction force. The force is represented as a uniformly distributed traction acting over a characteristic microstructural length, $2d$ [Fig. 3(c)]. This geometric choice allows the effect of contact to be expressed in terms of normal and shear forces, P and Q, respectively, acting over $2d$

$$P = F\sin\omega, \quad Q = F\cos\omega \tag{3a}$$

or

$$Q + iP = Fe^{i\omega}. \tag{3b}$$

The uniformly distributed normal force causes a crack opening, u, at $x = -d$

$$u = [4(1-v)/\pi G]P\chi(c) \tag{4a}$$

where [8]

$$\chi(c) = (2c)^{-1}\left\{\begin{array}{l}\sqrt{1+c}-\sqrt{1-c}+(1+c/2)\ln \\ \times [(\sqrt{1+c}+1)(\sqrt{1-c}-1)^{-1}] \\ -(1-c/2)\ln \\ \times [(1+\sqrt{1-c})(1-\sqrt{1-c})^{-1}]\end{array}\right\}$$

while the uniformly distributed shear force causes a relative shear displacement, v; at $x = -d$

$$v = \frac{4(1-v)}{\pi G}Q\chi(c) \tag{4b}$$

where G is the shear modulus and v is Poisson's ratio. The crack surface displacements caused by contact are related to the forces by

$$u + iv = \frac{4(1-v)}{\pi G}\chi(c)Fe^{i\omega} \tag{5}$$

such that the contact induced contribution to the stress intensities at the crack tip, K', are

$$K'_{II} + iK'_{I} = \sqrt{2/\pi d}f(c)Fe^{i\omega} \tag{6}$$

where

$$f(c) = (\sqrt{1+c}-\sqrt{1-c})/c$$

and the subscripts I and II refer to the opening and shear modes, respectively.

The corresponding crack surface displacements at the contact site ($x = -d$) dictated by the remote loads are

$$v + iu = [4(1-v)/G]\sqrt{d/2\pi}(K_{II} + iK_{I}) \tag{7}$$

where K_{I} and K_{II} are the stress intensities associated with the applied loads.

3. CONTACT AND LOCKING CONDITIONS

The existence of crack surface contact and of frictional locking are influenced by the relative magnitudes of the phase angle of loading, Ψ, the friction angle, ϕ, and the kink angle, β. Contact along the

that only occurs when

$$u\cos\beta + v\sin\beta > 0. \tag{8}$$

Hence, by defining a characteristic stress intensity K^* as

$$K^* = -(K_{II}\cos\beta + K_{I}\sin\beta) \tag{9}$$

contact of the facet faces occurs when

$$K^* > 0 \tag{10a}$$

which, since $\tan\Psi = K_{II}/K_{I}$, is equivalent to

$$\pi - \beta < \Psi < 2\pi - \beta. \tag{10b}$$

Furthermore, contact along the straight crack surfaces occurs when

$$K_{I} < 0. \tag{11}$$

The basic contact conditions are mapped in Fig. 4. For present purposes, the condition, $K_{I} < 0$, is not meaningful. The analysis should only be used for positive K_{I}. A separate analysis would be needed for negative Mode I.

Frictional locking occurs when the tangential force along the kink, T, satisfies the inequality

$$T < \mu N \tag{12}$$

where N is the normal force. This condition can be expressed in terms of the forces P and Q using

$$N = Q\cos\beta + P\sin\beta = \text{Re}[(Q + iP)e^{-i\beta}] \tag{13a}$$

$$T = Q\sin\beta - P\cos\beta = -\text{Im}[(Q + iP)e^{-i\beta}]. \tag{13b}$$

Then, by noting that

$$N\sin\phi - T\cos\phi = \text{Im}[(N - iT)e^{-i\phi}]$$

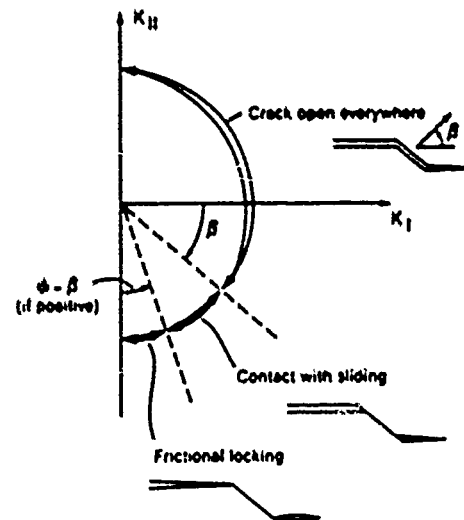$$= \text{Im}[(Q + iP)e^{-i\phi}e^{-i\beta}] \tag{14}$$

Fig. 4. A map of the sliding and locking conditions anticipated with the crack asperity. For present purposes, only $K_{I} > 0$ is relevant.

the locking condition becomes

$$\text{Im}[(Q + iP)e^{i\bar{z} \cdot K}] > 0 \qquad (15a)$$

or, because locking requires that $u = v = 0$, (3b), (5) and (7) give

$$\text{Im}[(K_{II} + iK_{I})e^{i\bar{z} \cdot h}] < 0. \qquad (15b)$$

Consequently, since

$$K_{II} + iK_{I} = |K|e^{-i\bar{z} \cdot \bar{z}} \qquad (16)$$

locking occurs when

$$\sin(\psi + \beta - \phi - \pi/2) > 0 \qquad (17a)$$

or

$$\pi/2 - \beta + \phi < \psi < 3\pi/2 - \beta + \phi \qquad (17b)$$

Thus, if $\phi - \beta > 0$, a range of loading exists in which $K_I > 0$ and the facets are frictionally locked as mapped in Fig. 4.

4. INTERFACE FRACTURE RESISTANCE

The above conditions of contact and frictional locking provide insights concerning the analysis of the interface fracture resistance. For purposes of analysis, the spectrum of contacts must be simulated by means of a contact model. Two such models are presented, each providing a different perspective of the contact phenomenon. One model considers a single row of contacts with friction having a full spectrum of facet angles located at a *fixed* distance from the crack front. The fracture resistance can then be estimated, in principle, by summing over the number of rows within the contact zone. However, to accomplish this, an estimate of the zone size and of interaction between rows is needed. For this reason, a second zone model is developed that examines a simple contact condition at the facets, but more rigorously addresses the zone size and interactions that occur along the zone.

4.1. Single row model

Contact behavior at facets within a row is governed by the net displacements of the facet surfaces, derived from equations (5) and (7) as

$$(u + iw)e^{-i\phi} = 4\frac{1-v}{G}$$

$$\times \left[\sqrt{\frac{d}{2\pi}} (K_{II} + iK_{I})e^{-i\phi} + \frac{g(t)}{\pi} Fe^{-i(t-w)} \right]. \qquad (18)$$

When sliding contact occurs, $\beta - \omega = \phi$ and $u \cdot n = 0$ or, equivalently, $\text{Re}[u + iw]e^{-i\phi}] = 0$. Thus, by (18), the resultant force across the facet faces is

$$F = \sqrt{\pi d/2} \, K^{\bullet}/\cos \phi. \qquad (19)$$

Inserting this expression for the force F into equation (6) and adding the stress intensities from the applied

loads gives

$$K'_{II} + iK'_{I} = K_{II} + iK_{I} + h(t)e^{-i\bar{z}}K^{\bullet}/\cos\phi \qquad (20)$$

where $h(t) = f(t)g(t)$ is the function plotted in Fig. 5. Consequently, the tip stress intensities are

$$K'_I = K_I + h(t)\sin(\beta - \phi)K^{\bullet}/\cos\phi$$

$$K'_{II} = K_{II} + h(t)\cos(\beta - \phi)K^{\bullet}/\cos\phi \qquad (21)$$

and the strain energy release rate at the tip is

$$G' = [(1 - v)/2G][(K'_I)^2 + (K'_{II})^2]. \qquad (22)$$

Noting that the effective energy release rate associated with the applied loads is

$$G = [(1 - v)/2G][K_I^2 + K_{II}^2] \qquad (23)$$

the crack tip energy release rate becomes

$$G' = G + \frac{1-v}{2G}$$

$$\times \left\{ \frac{2hK^{\bullet}[K_I\sin(\beta - \phi) + K_{II}\cos(\beta - \phi)]}{\cos\phi} \right.$$

$$\left. + \frac{h^2K^{\bullet 2}}{\cos^2\phi} \right\} \qquad (24)$$

Thus, by (9)

$$\frac{\Delta G}{G} = \Sigma(\phi, \beta, v, t)$$

$$= + 2h \frac{[(\sin\beta + \cos\beta\tan\psi)(\sin(\beta - \phi) + \cos(\beta - \phi)\tan\psi)]}{\cos\phi(1 + \tan^2\psi)}$$

$$- \frac{h^2(\sin\beta + \cos\beta\tan\psi)^2}{\cos^2\phi(1 + \tan^2\psi)} \qquad (25)$$

where $\Delta G = G - G'$ is the *reduction* in G, i.e. the shielding, caused by contact at the facet. (Recall that this equation, as well as all others above in-
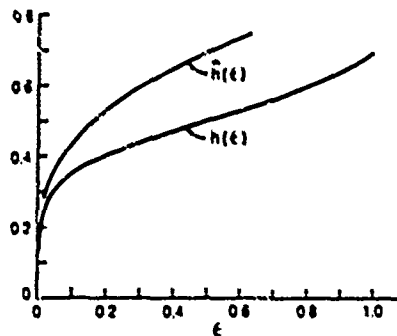

Fig. 5 Comparison of the functions $h(t)$ from the present approximate model and $\hat{h}(t)$ from the exact solution for a microcrack ahead of a macrocrack.
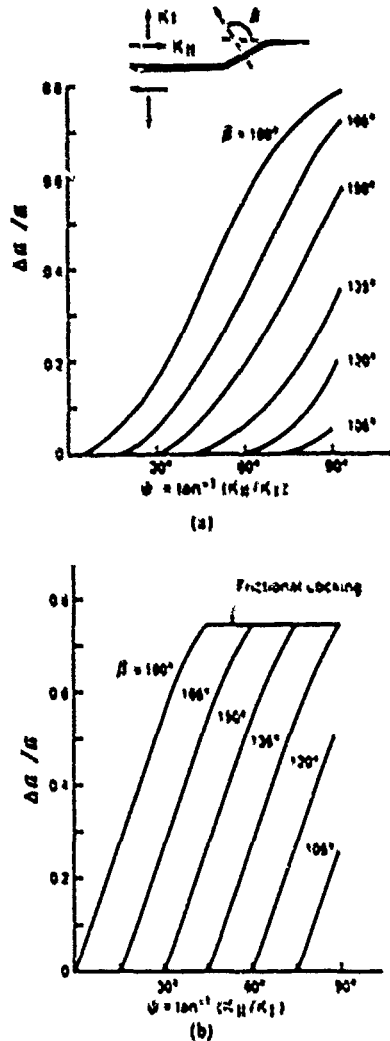
(a)

(b)

Fig. 6. (a) Reduction in crack tip release rate as a function of phase angle of loading for various kink orientations in the absence of friction ($h = 0.5$, $\phi = 0$). (b) Effect of friction ($\psi = 45°$) on reduction in near tip energy release rate ($h = 0.5$).

volving ϕ, are limited to $0 < \beta < \pi/2$. For the range $\pi/2 < \beta < \pi$, ϕ must be replaced by $-\phi$.)

When *locking* occurs [equation (17), Fig. 4], the model gives

$$K'_1/K_1 = K'_{11}/K_{11} = 1 - h(\epsilon)   (26)$$

and

$$G'/G = [1 - h(\epsilon)]^2 = \Omega(\epsilon)   (27)$$

which coincides with (21) and (24) when $\psi \to 3\pi/2 + \phi - \beta$. When locking occurs, the material along the kink behaves elastically and the tip stress intensities are equivalent to those for a microcrack of length $(1 - r)d$ at distance $2\epsilon d$ from the macrocrack tip

(Fig. 3(c)). The solution to this problem is known exactly and thus provides some calibration of the model. The solution is [9]

$$K'_1/K_1 = K'_{11}/K_{11} = 1 - \hat{h}(\epsilon)   (28)$$

where $\hat{h}(\epsilon)$ is plotted in Fig. 5 along with $h(\epsilon)$. Noting the roles of h in (26) and $\hat{h}$ in (28), it is concluded that the present model underestimates somewhat the reduction in crack tip intensities due to contact at the facet. In the results calculated below, values of h are prescribed (rather than ϵ). To improve the accuracy of the model, these values may be identified with $\hat{h}$ for the purpose of evaluating the associated relative facet width ϵ.

Curves of $\Delta G/G$ vs ψ for various β as calculated from (25) are plotted in Fig. 6(a) with $\phi = 0$ and in Fig. 6(b) with $\phi = \pi/4$, in each case with $h = 0.5$. For loading combinations in the range shown ($0 < \psi < \pi/2$), contact does not occur for facet orientations with β less than $\pi/2$. The effect of friction is greatest when the facet angle is large, i.e. β greater than about $3\pi/4$. For the example in Fig. 6(b), frictional locking only occurs for $\beta > 3\pi/4$.

The interface fracture resistance is estimated based on consideration of a row of contacts with a uniform distribution of kink angles parallel to the front, ranging from 0 to π. (Kink angles with $\beta < 0$ or $\beta > \pi$ are unlikely because alternative, noninterfacial, fracture paths would normally obtain at facets having $\beta < 0$ or $\beta > \pi$.) For loading combinations with $0 \leqslant \psi \leqslant \pi/2$, the following conditions pertain. No contact occurs if

$$0 < \beta < \pi - \psi.   (29)$$

If $\phi + \psi < \pi/2$, no locking occurs at any β and the range of sliding contact is

$$\pi - \psi < \beta < \pi.   (30a)$$

If $\phi + \psi > \pi/2$, sliding contact occurs for

$$\pi - \psi < \beta < 3\pi/2 - \phi - \psi   (30b)$$

while locking occurs for

$$3\pi/2 - \phi - \psi < \beta < \pi.   (30c)$$

The *local* values of $\Delta G/G$ along the crack front are: zero for regions of no contact, Σ for contact without locking [equation (25)], and $1 - \Omega$ for locked kinks [equation (27)]. Interaction effects caused by neighboring kinks having different contact/locking conditions are thus neglected. With this simplification, an average value of ΔG along the crack front can be obtained by integration. For the case $\phi + \psi < \pi/2$, the fraction of noncontacting kinks is $1 - \psi/\pi$, while the fraction of contacting, non-locking kinks is ψ/π. Thus

$$\langle \Delta G \rangle/G = \frac{1}{\pi} \int_{\pi-\psi}^{\pi} \Sigma \, d\beta.   (31a)$$
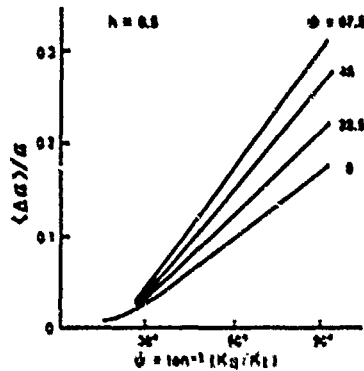
Fig. 7 Variation in crack shielding ⟨ΔG⟩ with phase angle of loading caused by a row of contacts, for several friction angles (h = 0.5).

Similarly, for $\phi + \psi > \pi/2$

$$\langle \Delta G \rangle / G = \frac{1}{\pi} \int_{\phi - \psi}^{3\pi/2 - \psi} \Sigma \, d\beta$$

$$+ \left(\frac{\phi + \psi}{\pi} - \frac{1}{2} \right) (1 - \Omega). \quad (31b)$$

Curves of ⟨ΔG⟩/G as a function of ψ obtained from (31a, b) are plotted in Fig. 7 for several values of the friction angle φ. These curves were obtained using h = 0.5. The effect of different choices for h ranging over all realistic possibilities (cf. Fig. 5) is shown in Fig. 8. These results can be used to illustrate some of the issues involved in determining the extent of crack shielding.

It is firstly evident that there is only a moderate effect of the friction coefficient on crack shielding, within the usual range $0 < \phi < \pi.4$, because locking constitutes the greatest impediment to crack surface displacement. It is also noted from Fig. 5 that h remains quite large (≳ 0.2) down to quite small values of ε(~ 0.05). Consequently, by simply sum-

ming ⟨ΔG⟩ on parallel rows of nominally identical facets, all governed by equation (31), the net shielding must be sensitive to the number of rows and hence, the size of the contact zone. These insights suggest that a zone model which emphasizes the locking characteristics and explicitly incorporates the zone size should provide a more reasonable prediction of the effects of contact on the interface fracture resistance. Such a model is presented below.

4.2. Zone model

A zone model is developed for the simplified contact conditions depicted in Fig. 9 corresponding to $\beta = 0$ and $\phi = 0$ (no friction). Then, $K_I^t = K_t$ and the crack opening depends on K_I only, as given by

$$u(r) = 8(1 - \nu^2)K_{IV} \, r \sqrt{2\pi E} \quad (32)$$

where r is the distance from the crack tip. Contacts exist over a zone length, L, that satisfies the condition

$$u(L) = H$$

where H is the height of the interface step (Fig. 9), such that

$$L = (\pi \, 32)[EH \cdot (1 - \nu^2)K_I]^2 \quad (33)$$

Within the contact zone, the shear stresses and displacements are elastic and analogous to those associated with a linear array of microcracks (Fig. 9). Furthermore, in order to evaluate K_{II}, the microcracks can be simulated by a continuous linear spring model (Fig. 9) in which the stresses τ and
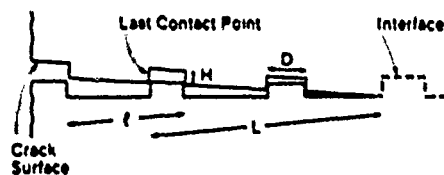
a) Zone Configuration
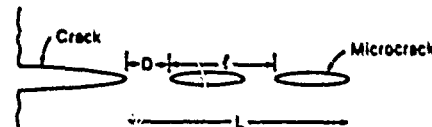
b) Microcrack Zone

c) Elastic Spring Configuration
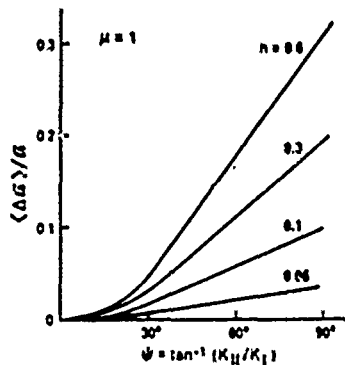

Fig. 8 The variation in crack shielding ⟨ΔG⟩ with phase angle caused by a row of contacts, for a range of $h(\mu = \tan \phi = 1)$.

Fig. 9. A schematic illustrating the zone model used to determine trends in G_i with phase angle of loading.

Table 1 Interface properties for the
glass polymer system

G_a	4 Jm
E	70 GPa
H	10^{-6} m
$t \approx 0.1 L/H$ G_a $\approx 10^3$	

displacements r are related by

$$ r = \frac{S t/(1 - v^2) \ln[1 \sin(\pi D/2l)]}{\pi E} \quad (34) $$

where l is the spacing between facet (microcrack) centers and D is the facet length (Fig. 9). This is the exact result for the additional shear displacement due to the presence of an infinite linear array of microcracks subject to remote shear stress τ parallel to the cracks [11, 12]. The linear spring model has been solved by Budiansky et al [13]. For the present linear spring (34), the result of interest is

$$ K'_{11}/K_{11} = k(z) \quad (35) $$

where

$$ z = \frac{(L/l)}{\ln[1 \cdot \sin(\pi D/2l)]} \quad (36) $$

and the function $k(z)$ is given in Table 1 (as $1/\kappa$) in Ref [13]. Plots of K'_{11}/K_{11} vs D/l from (35) are shown in Fig. 10 for various numbers of microcracks $N(L = Nl)$. The result for one microcrack is the exact result from Ref. [9].

Using K'_{11}/K_1 and (35), the relation between energy release rates is obtained as

$$ G'/G = \frac{1 + k^2(z) \tan^2 \psi}{1 + \tan^2 \psi}. \quad (37) $$

Further progress is achieved by combining equations (33) and (36) to give

$$ z = \frac{\pi E H^2 (G'/G)(1 + \tan^2 \psi)}{32(1 - v^2)/G' \ln[1 \cdot \sin(\pi D/2l)]}. \quad (38) $$

The basic trends in the fracture behavior of the interface with the phase angle of loading can be

estimated by selecting the value of z at $\psi = 0$ as a reference value. z_a, and setting G' equal to the fracture resistance of the interface G_a, such that

$$ z_a = \frac{\pi (E H^2 / G_a)}{32(1 - v^2) \ln[1 \sin(\pi D/2l)]}. \quad (39) $$

The quantity z_a contains the basic information concerning the interface and is thus a *material parameter*. The results contained in equations (35), (37), (38) and (39) can be combined to provide an expression for the crack shielding as

$$ \Delta G/G = \frac{\tan^2 \psi \{1 - k[z_a(1 + \tan^2 \psi)](\Delta G/G + 1)\}}{1 + \tan^2 \psi}. \quad (40) $$

Specific trends in $\Delta G/G$ with ψ for various z_a determined using equation (40) are plotted in Fig 11

It may be ascertained from Fig 11 that two regimes exist, one at $z_a \gtrsim 1$, and the other when $z_a \lesssim 10^{-1}$. For the former, contact has the maximum effect on crack shielding, such that, $K'_{11} \approx 0$ and

$$ \Delta G/G = \frac{\tan^2 \psi}{1 + \tan^2 \psi}. \quad (41) $$

For the latter, there is essentially no shielding when $\psi \lesssim \pi/2$. The significance of these two regimes may be appreciated by noting that undulating interfaces typically have geometry $D/l \sim 1/2$, $H/l \sim 1/2$. Then z_a becomes

$$ z_a \approx 0.1(E H/G_a)(1 - v^2). \quad (42) $$

The governing material parameter is thus

$$ \chi = E H/G_a. $$

This parameter, in essence, determines the contact zone dimensions, such that small values result in no contact and large values give full contact. Furthermore, the transition between these extremes occurs over a relatively small range of χ between ~ 10 and $\sim 10^{-2}$. Consequently, either a small value of the intrinsic interface toughness G_a or a large amplitude
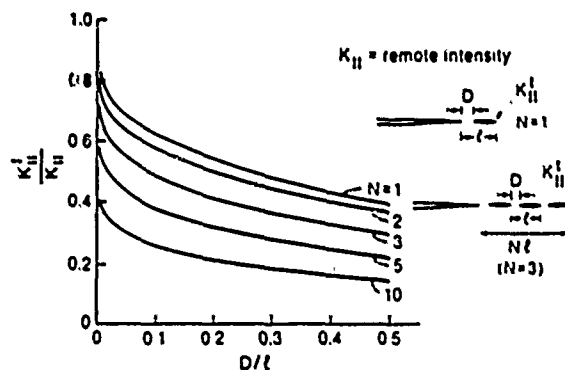

Fig. 10. Stress intensity at tip of lead microcrack in an array of equally spaced microcracks ahead of a macrocrack.
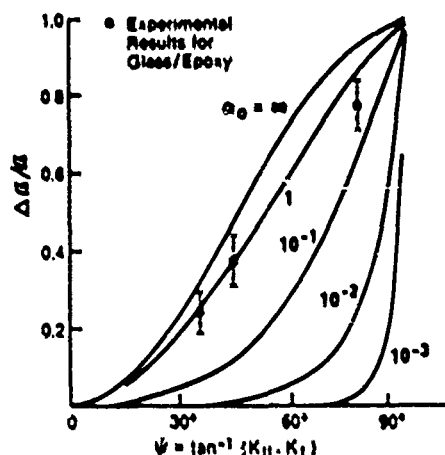
Fig. 11. Variation in crack shielding ΔG with phase angle of loading for various values of the contact zone parameter, x. Also shown are some experimental results obtained for a glass/polymer interface [10].

undulation, H, may cause χ to become large, resulting in a mixed mode fracture resistance governed by equation (41), such that

$$G_i = G_i(1 + \tan^2 \psi). \tag{43}$$

Conversely, when H is small and/or G_0 is large, contact does not occur and $G_i \approx G_0$ for $\psi < \pi/2$.

5. COMPARISON WITH EXPERIMENT

Experimental results have been obtained in separate studies [10]. The results are in general accordance with the zone model predictions as depicted on Fig. 11 with $x_0 \approx 1$ (or $\chi \approx 10$). Values of χ independently determined for the test interface are somewhat larger ($\chi \sim 10^2$, Table 1). This difference probably derives from the geometric simplicity of the zone model (Fig. 9). In particular, more realistic interface geometries would allow sliding of the edge of the contact zone and reduce the effective magnitude of the contact tractions, leading to improved correlations between theory and experiment. Some aspects of the discrepancy may also be associated

with the modulus mismatch across the interface, which is not explicitly considered in the present analysis.

6. CONCLUDING REMARKS

The present analysis of contact effects on interface fracture resistance indicates that a simple zone model without friction predicts trends with the phase angle of loading qualitatively consistent with experimental results for a brittle interface that has no obvious plasticity associated with crack propagation. More complete experimental studies of contact zones would evidently allow further progress, and highlight deficiencies associated with the neglect of friction and with the present geometric simplification used to describe the interface. Studies of trends with the zone size parameter, $EH \cdot G_0$, would be particularly insightful in this regard.

When the interface is subject to normal compression (negative K_i'), a different analysis of fracture is needed. Nevertheless, it is already evident from the present analysis that sliding induces a positive K_i' at the tip over a substantial range of compressions. Consequently, crack propagation will still be possible, as observed experimentally [10].

REFERENCES

1. A. G. Evans and J. W. Hutchinson. Int. J. Solids Struct. 24, 455 (1984).
2. M. S. Hu, M. D. Thouless and A. G. Evans, Acta metall. 36, 1301 (1988).
3. P. G. Charalambides, J. Lund, R. M. McMeeking and A. G. Evans, J. appl. Mech. In press.
4. P. G. Charalambides and A. G. Evans, J. Am. Ceram. Soc. To be published.
5. M. Rühle, unpublished research.
6. M. D. Drory, M. D. Thouless and A. G. Evans, Acta metall. 36, 2019 (1988).
7. J. R. Rice, J. appl. Mech. 35, 379 (1968).
8. H. Tada, P. Paris, G. Irwin, The Stress Analysis of Cracks Handbook. DEL Research Corp. (1973).
9. F. Rose, Int. J. Fract. 31, 233 (1986).
10. H. C. Cao, Ph.D. thesis, Univ. of California, Berkeley (1988); H. C. Cao and A. G. Evans, Mech. Mater. In press.
11. W. T. Koiter, An infinite row of collinear cracks in an infinite elastic sheet, Ingen. Arch. 28, 168 (1959).
12. M. Ortiz, Int. J. Solids Struct. 24, 231 (1988).
13. B. Budiansky, J. C. Amazigo and A. G. Evans, J. Mech. Phys. Solids. 36, 167 (1988).

MATERIALS
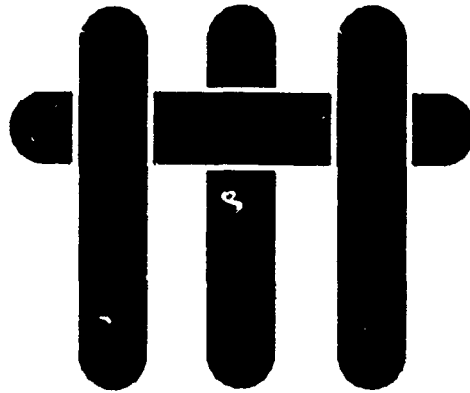
ON CRACK PATH SELECTION AND THE INTERFACE FRACTURE ENERGY IN BIMATERIAL SYSTEMS

by

A. G. Evans, B. J. Dalgleish, M. He
Materials Department
College of Engineering
University of California
Santa Barbara, California 93106

and

J. W. Hutchinson
Division of Applied Sciences
Harvard University
Cambridge, Massachusetts

ABSTRACT

The intent of this article is to apply recent solutions for the mechanics of cracks at and near bimaterial interfaces to rationalize crack trajectories observed by experiment and to provide a basis for interpreting measurements of the interface fracture energy, Γ_i. It is demonstrated that the choice of test specimen governs the tendency of cracks to either remain at interfaces or deviate away, based on considerations of the phase angle of loading, ψ. It is further revealed that the measured interface fracture energy may be strongly influenced by the crack trajectory, as governed by ψ, through crack shielding and plasticity effects. Consequently, interfaces do not typically have unique fracture energies, but instead Γ_i depends on ψ which, in turn, is influenced by the test method.

1. INTRODUCTION

Fracture along and adjacent to bimaterial interfaces has several morphological manifestations. In some cases, the fracture occurs at the interface, while in others fracture occurs in the more brittle constituent[1,2]. Furthermore, cases exist wherein the fracture alternates between the interface and the adjacent material (Fig. 1) and in other instances, the crack deviates from one interface to the other[1,4] (Fig. 2). The path selected by the crack also influences the fracture energy. In particular, fracture energies Γ_i have bee.. measured for "interface" cracks wherein Γ_i has substantially exceeded that for the brittle constituent, Γ_s.[4,5,6] This behavior is seemingly at variance with intuitive reasoning that cracks should propagate in the "material" having the lowest fracture energy. The present article constitutes an attempt to rationalize these features of interface fracture by invoking recent mechanics solutions for cracks at and near interfaces.[7]

2. CRACK TRAJECTORIES

Various hypotheses have been made concerning crack trajectories in homogeneous brittle solids. Experiments that exhibit sensitivity to the choice of crack path criterion involve the decohesion of thin films from brittle substrates,[8-10] particularly the location of the steady-state crack path within the substrate[11] (Fig. 3). The experimental results are consistent with the postulate that the crack acquires a trajectory in which the mode II stress intensity factor, K_{II} is zero. A corollary of this criterion is that a crack cannot remain in a plane having non-zero K_{II}, but instead is deflected in a direction that reduces K_{II} toward zero.

The above crack path criterion clearly does not apply when the crack is at an interface, because consideration of the relative fracture energy, Γ_i/Γ_s, is also

3

involved. Analysis of the kinking of a crack out of an interface between two brittle solids[7] has revealed that the crack path depends both on Γ_i/Γ_s and the relative shear v to opening, u, experienced by the interface crack, as characterized by the phase angle ψ[12]

$$\psi = \tan^{-1}(v/u) - \epsilon \ell nr - \tan^{-1}2\epsilon \tag{1a}$$

where

$$\epsilon = \frac{1}{2\pi} \ell n\left(\frac{1-\beta}{1+\beta}\right) \tag{1b}$$

with β being one for the Dundurs' parameters, α and β;

$$\alpha = \frac{\mu_1(1-v_2) - \mu_2(1-v_1)}{[\mu_1(1-v_2) + \mu_2(1-v_1)]}$$

$$\beta = \frac{\mu_1(1-2v_2) - \mu_2(1-2v_1)}{2[\mu_1(1-v_2) + \mu_2(1-v_1)]} \tag{1c}$$

and r is the distance from the crack tip. Note that, for $\beta = 0$ (typical of many systems of interest),

$$\psi = \tan^{-1}(v/u) = \tan^{-1}(K_{II}/K_I) \tag{2}$$

and thus ψ is a direct measure of the relative mode II loading on the interface crack. Trends in the incidence of kinking out of the interface with Γ_i/Γ_s and ψ are shown in Fig. 4, for the range in α of practical interest.

4

The solutions of Fig. 4 may be interpreted by equating Γ_s to the fracture energy of material 2 such that the sign of α is consistent with Eqn. (1c). It is apparent that the preferred path of the crack is influenced by the magnitude of the phase angle, such that cracking out of the interface is most likely when $\psi = 70°$ and, furthermore, that the crack prefers to extend into lower modulus material (subject to ψ having the appropriate sign, i.e., material 2 in Fig. 4 being the lower modulus member).

A related problem of substantial importance concerns the case wherein one of the materials is ductile and the other is brittle. Then the fracture behavior and the "interface" fracture energy are extremely sensitive to the sign of the phase angle. Some of the basic behaviors are summarized in Fig. 5. When the phase angle has a positive sign, in accordance with the sign convention given in Fig. 5, and material S is the brittle member (such that $\Gamma_s < < \Gamma_f$), the relative incidence of cracking out of the interface is identical to that described for the all-brittle system (Fig. 4). However, when the phase angle is negative, the large fracture energy of material F, compared with the interface ($\Gamma_f > > \Gamma_i$) prohibits propagation of the crack out of the interface. Then, one of two possibilities occurs, depending upon the yield strength of the ductile member. For a low yield strength material, plastic blunting of the interface crack is observed (Fig. 6a) and failure often occurs by ductile mechanisms (Fig. 6b) involving hole nucleation at the interface.[3] Alternatively, the stress field of the interface crack interacts with preexisting flaws in the brittle material, causing cracks to grow from these flaws back toward the interface, resulting in a serrated fracture with "chips" of brittle material attached to the fracture surface (Figs. 1 and 5). This behavior can be understood by noting that the flaws in the brittle material are subject to substantial mode II loading having sign that leads to crack extension toward the interface.

3. THE INTERFACE FRACTURE ENERGY

The above discussion concerning crack paths has important implications for measuring and interpreting information relating to the "fracture energy of interfaces." Most importantly, the test configuration determines the sign and the magnitude of ψ, which in turn governs the crack path and thus, the mechanisms that contribute to Γ_i. Several examples are used to illustrate the salient features: "mode I" sandwich specimens[4,13] (Fig. 7a), peel tests[5] (Fig. 7b) and film decohesion tests[14] (Fig. 7c). In all cases, residual stress contributes substantially to the observed behavior and in the interpretation of the results.

3.1 SANDWICH TESTS

The interpretation of fracture tests performed on sandwich specimens is relatively straightforward when the precrack length, a, is very large compared with the bond layer thickness, h, (a $\gtrsim$ 30h). Then, the residual strain does not contribute to the energy release rate $\mathcal{G}$, because the bond layer remains attached to one side of the specimen and thus, retains its state of residual strain.[15] When this crack length condition is satisfied, the only other significant consideration is the influence of the bond layer on the magnitude of $\mathcal{G}$ and ψ. When h is small compared with the specimen thickness, d, the energy release rate is unaffected by the layer.[13] However, the presence of the layer causes ψ to deviate from zero. For the typical case, $\beta \approx 0$, ψ is governed solely by α, with the trend[13] plotted on Fig. 8. While the range in ψ is small, it is often sufficient to influence the crack trajectory. Specifically, in most cases, the shear modulus of the bond layer is less than that of the bonded member, whereupon the sign of ψ directs the crack toward the interface. Consequently, when the bond material is ductile, the crack tends to either remain at the interface or

follow a serrated path near the interface. In either case, plasticity in the bond layer can contribute to Γ_i, causing the measured fracture energy to be substantially greater than Γ_s *even though the crack remains near the interface.*

When the crack is not long compared with the layer thickness, residual stress influences both K_I and K_{II} and must be taken into account.[15] Most significantly, when the bond layer has the larger thermal expansion coefficient, the sign of ψ tends to divert the crack away from the interface and furthermore, is sufficiently large ($\psi \sim 45°$)[15] that it can dominate the overall phase angle at the fracture criticality. In this case, cracks are often diverted into the brittle material, outside the bond.[2] The nominal fracture energy measured in such cases can have a broad range of values that depend on the crack length and the residual strain.[16] However, the actual fracture energy can be deduced if the residual strain contribution to G is properly taken into account.

Bond layers having a lower thermal expansion coefficient invariably generate a phase angle that diverts cracks toward the interface. Then, the measured fracture energy can either be larger or smaller than Γ_s, depending upon the residual strain contribution to G, the influence of plasticity, roughness, etc.

3.2   FILM DECOHESION

Films and coatings with interface edge cracks that cause decohesion experience mixed mode conditions, with ψ dependent largely on the sign of the residual stress.[13] Films in tension give $\psi \sim 45°$ with sign that tends to deflect the crack away from the interface. Consequently, when the substrate is brittle and the interface fracture energy is relatively high, the decohesion process proceeds in the substrate by cracking at a characteristic depth beneath the interface.[9-11] The substrate fracture energy Γ_s then becomes the relevant quantity. When the substrate is

ductile, the incidence of decohesion is governed by Γ_i, which may be influenced by plasticity in the substrate, etc. When decohesion does occur at the interface (Fig. 9), the decohesion analysis[11] may be used to measure Γ_i'.

Residual compression in the film subjects edge cracks to mode II conditions, plus compression normal to the interface. As yet, there is no understanding of interface fracture in such circumstances.

3.3   PEEL TESTS

The peel test is generally used for measuring the fracture energy of ductile compliant films on substrates.[5] The phase angle has a negative sign that causes substrate cracks to deviate into the interface. The test thus encourages the modes of interface fracture depicted on the left side of Fig. 6. The measured fracture energy can be appreciably less than the substrate value Γ_s and include contributions from crack tip plasticity in the film material. One complication with this test is that the peel load is sensitive to non-linear deformation *in the film*, such that deconvolution to obtain Γ_i requires knowledge of the plastic flow stress of the film.[5] However, when this effect is taken into account, valid Γ_i measurements are certainly possible and indeed, can be large compared with Γ_s.

4.   CONCLUDING REMARKS

A rational mechanics basis for predicting and interpreting crack trajectories in bimaterial systems has been proposed, based on knowledge of the phase angle of loading, ψ. In particular, the sign and magnitude of ψ dictate whether cracks either propagate along an interface or deviate into the adjoining material and extend

parallel to the interface. More importantly, phase angle conditions that motivate cracks to remain at the interface can, in some cases, yield interface fracture energies Γ_i in excess of that for the brittle member *even through the crack propagates at the interface*. Such high fracture energies contain contributions from plasticity, roughness-induced crack shielding, etc.

The phase angle of loading is strongly influenced by the choice of test specimen and thus, Γ_i experimentally determined on the same bimaterial interface can vary appreciably between test specimens. However, it is believed that a unique relationship exists between Γ_i and ψ for a given interface. It is thus emphasized that test specimens *must* be characterized *both* in terms of an *energy release rate* and *a phase angle of loading*. As more data on $\Gamma_i(\psi)$ are generated, the mechanisms that contribute to Γ_i can be addressed, and allow the development of a fundamental evaluation of interface fracture.

An important corollary of the above conclusion is that the prediction of such events as fiber debonding in composites, film decohesion and bond fracture require that *both* G *and* ψ be evaluated for the problem. Then, provided that Γ_i has been measured in the appropriate range of ψ, prediction can be made. This point is emphasized because ψ is typically much more difficult to calculate than G, and its evaluation is essential if interface cracking problems are to be adequately understood. Furthermore, schemes for calculating ψ using either integral equations or finite elements have been developed.

REFERENCES

[1]   B. J. Dalgleish, M. C. Lu and A. G. Evans, Acta Met. 36 (1988) 2029.

[2]   K. Burger, D. Brenner and G. Petzow, Z. Zahnärztl Implantol III (1987) 547.

[3]   B. J. Dalgleish, K. P. Trumble and A. G. Evans, Acta Met, in press.

[4]   T. S. Oh, J. Rödel, R. O. Ritchie and R. M. Cannon, Acta Met. 36 (1988) 2083.

[5]   K. S. Kim and N. Aravas, Intl. Jnl. Solids Structures 24 (1988) 417.

[6]   R. F. Pabst and G. Elssner, Jnl. Mater. Sci. 15 (1980) 188.

[7]   M. He and J. W. Hutchinson, Jnl. Appl. Mech, in press.

[8]   M. D. Thouless, A. G. Evans, M. F. Ashby and J. W. Hutchinson, Acta Met. 35 (1987) 1333.

[9]   M. S. Hu, M. D. Thouless and A. G. Evans, Acta Met. 36 (1988) 1301.

[10]   M. D. Drory and A. G. Evans, J. Am. Ceram. Soc., submitted.

[11]   Z. Suo and J. W. Hutchinson, Intl. Jnl. Frac., in press.

[12]   J. R. Rice, Jnl. Appl. Mech. 55 (1988) 98.

[13]   Z. Suo and J. W. Hutchinson, Matr. Sci. Eng., in press.

[14]   M. S. Hu and A. G. Evans, Acta Met. 37 (1989) 917.

[15]   H. C. Cao, M. D. Thouless and A. G. Evans, Acta Met. 36 (1988) 2037.

[16]   H. P. Kirchner, J. C. Conway and A. E. Segall, Jnl. Amer. Ceram. Soc. 70 (1987) 104.

FIGURE CAPTIONS

Fig. 1. Fracture of a Au/Al$_2$O$_3$ bond indicating fragment of Al$_2$O$_3$ attached to Au

Fig. 2. A crack in Al$_2$O$_3$ bonded with Au showing a crack alternating between interfaces

Fig. 3. Film decohesion by substrate cracking indicating the predicted crack location based on K$_{II}$ = 0 and experimental data obtained for two material combinations.

Fig. 4. A map representing the region of interface cracking for three values of the elastic mismatch parameter α. Positive ψ is defined by the mode II arrows in the inset diagram.

Fig. 5. The region of interface fracture for a bimaterial system when one material, F, has a much higher fracture energy than the other, S (Γ_f >> Γ_s). The result is plotted for the case, $\alpha = \beta = 0$. Note that when $\psi < 0$, fracture at the interface can incorporate segments of adjoining material detached at flaws.

Fig. 6. (a) Plastic blunting at the interface: a crack in the Al$_2$O$_3$/Al system
       (b) Ductile fracture surface: failure adjacent to the interface in the Al$_2$O$_3$/Al system

Fig. 7. A schematic indicating the various tests used to measure the interface fracture energy

Fig. 8. The phase angle shift ω for a sandwich specimen as a function of the elastic mismatch across the interface, expressed through the Dundurs' parameter α

Fig. 9. Film cracking followed by interface decohesion for Cr films deposited onto stainless steel

11

FIG. 1

460µm

FIG. 2

FIG. 3

FIG. 4

FIG. 5

alumina

aluminum

FIG. 6A

ALLOY SIDE

INTERFACE SIDE

HOLE
NUCLEATION
AT
INTERFACE
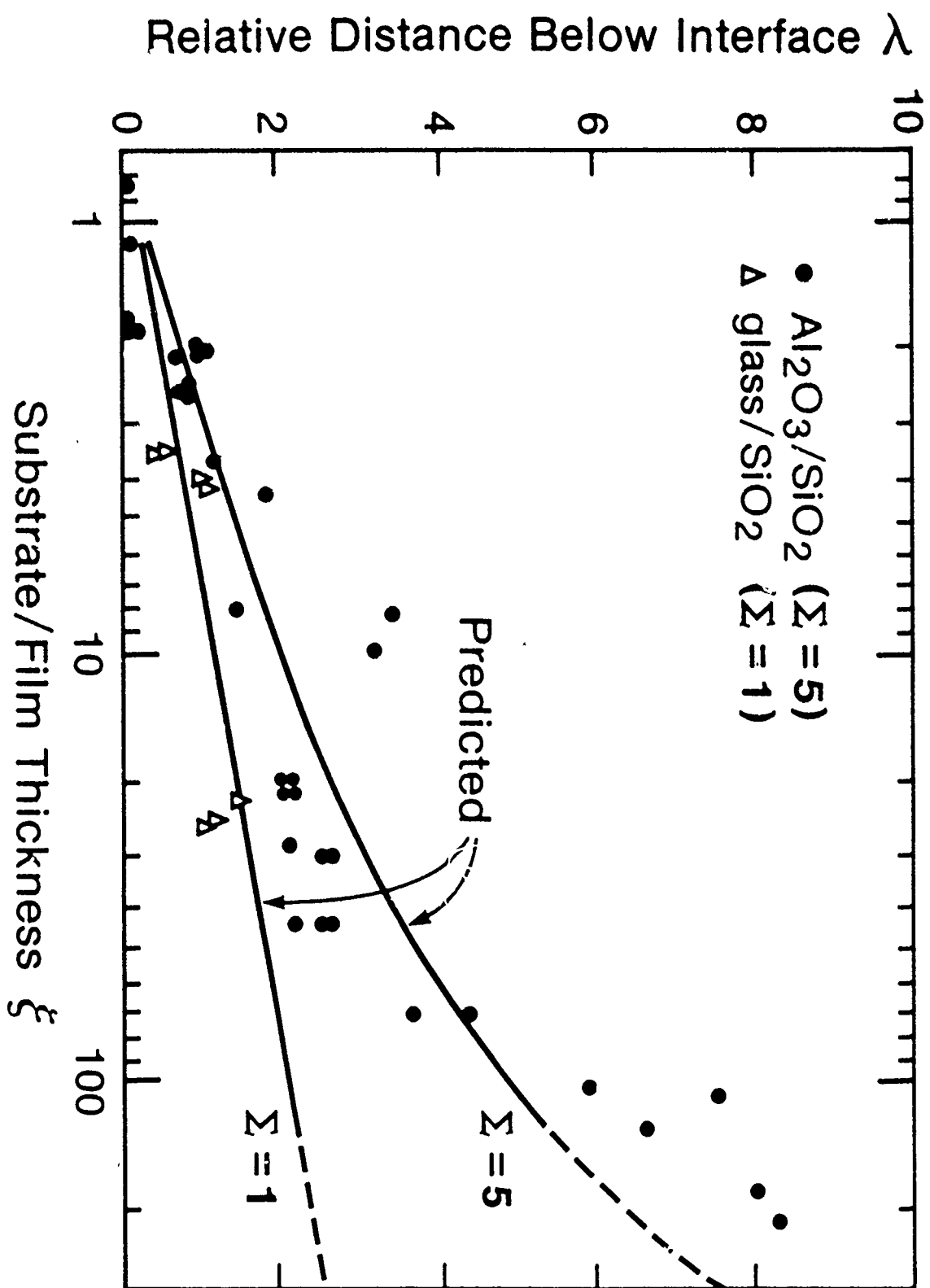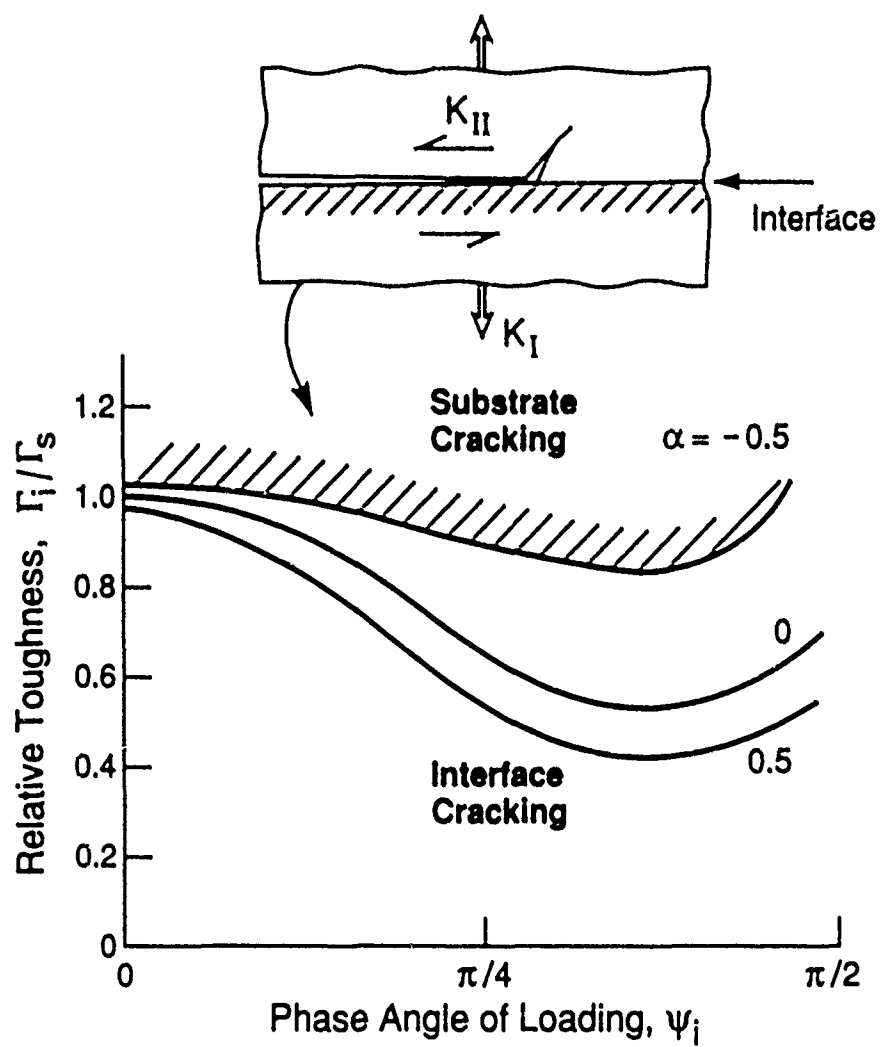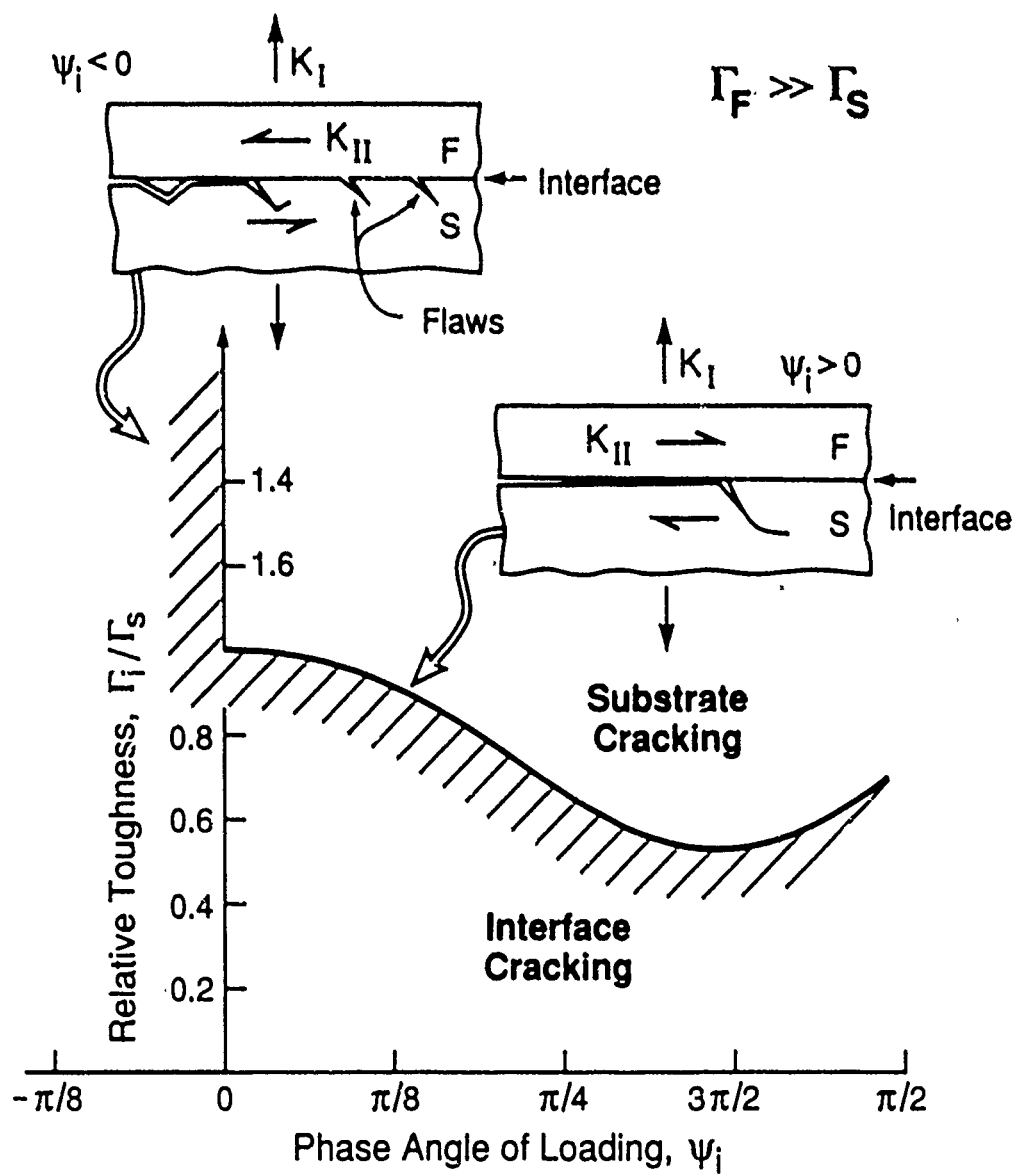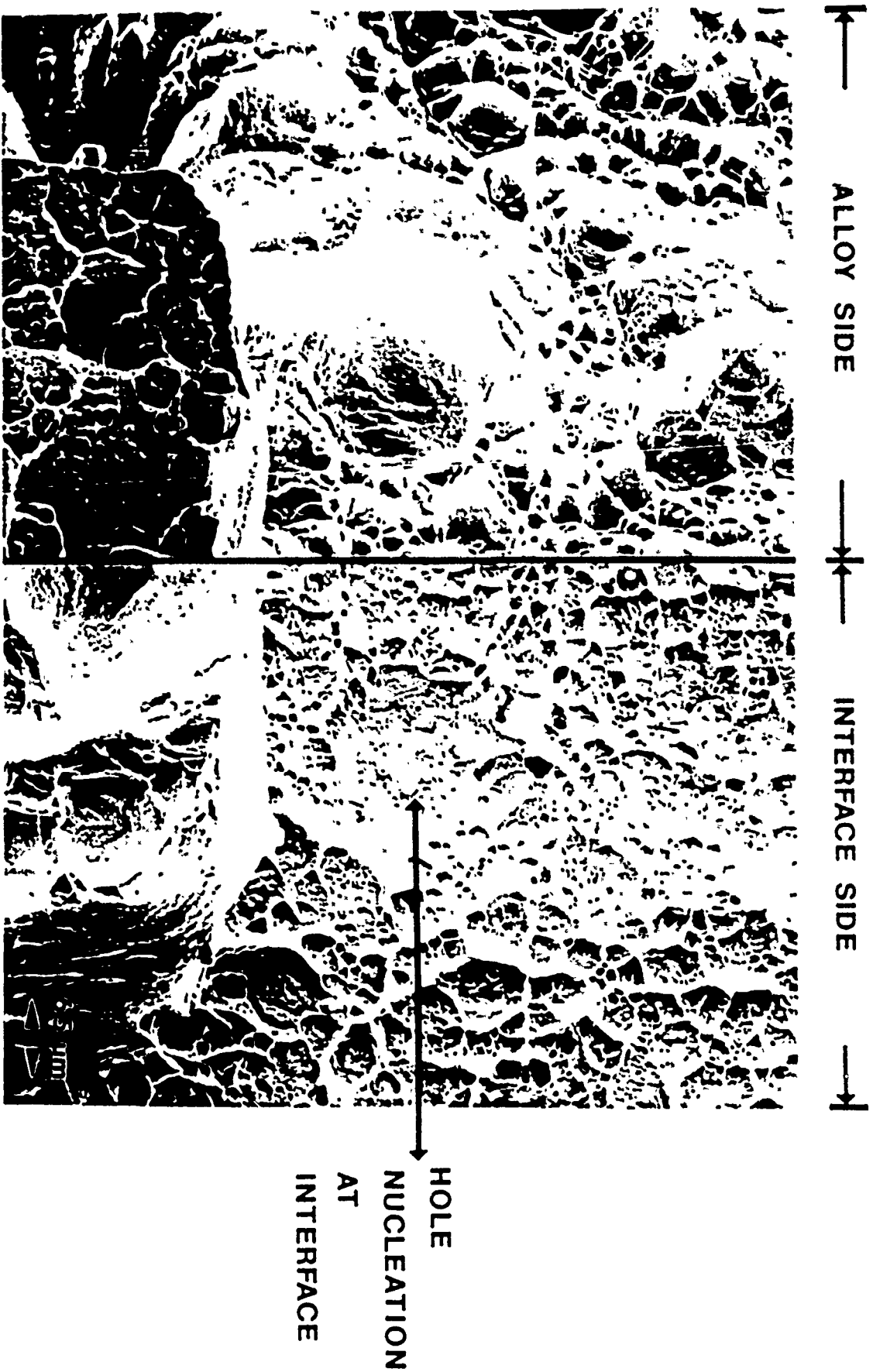
FIG. 6B

$b^2 h^3 E\mathcal{G}/P^2$
$= 12 (a/h)^2 [1+(2/3) (h/a)]^2$
$0 \lesssim \psi \lesssim 5°$
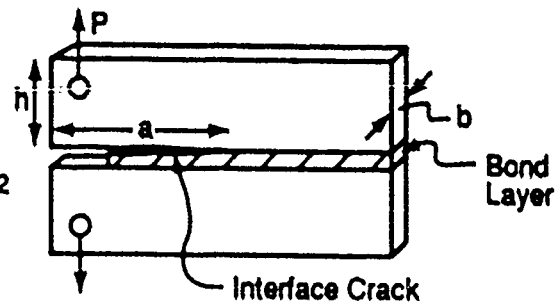
a) Sandwich Test Specimen

$\mathcal{G}$ depends on non-linear properties of film
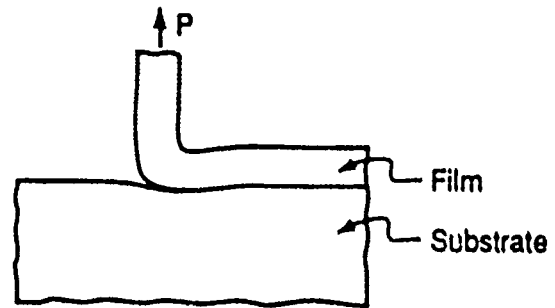
b) Peel Test

$b^2 H^3 E\mathcal{G}/P^2 \ell^2$
$= k (h/H)$
$\psi \approx 50°$

c) Decohesion Test

$M = P\ell/2b$
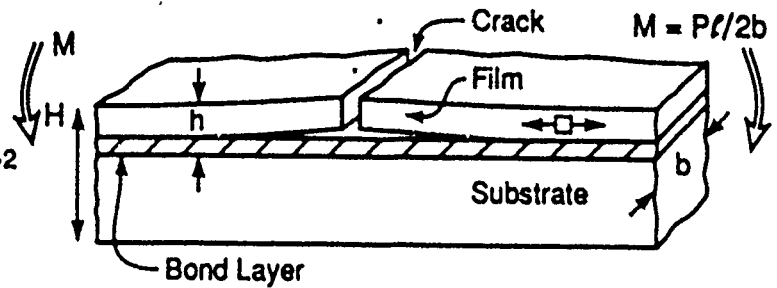
$\mathcal{G}E/\sigma^2 R \approx 0.25$
$70° \lesssim \psi \lesssim 90°$

d) Composite Cylinder Test
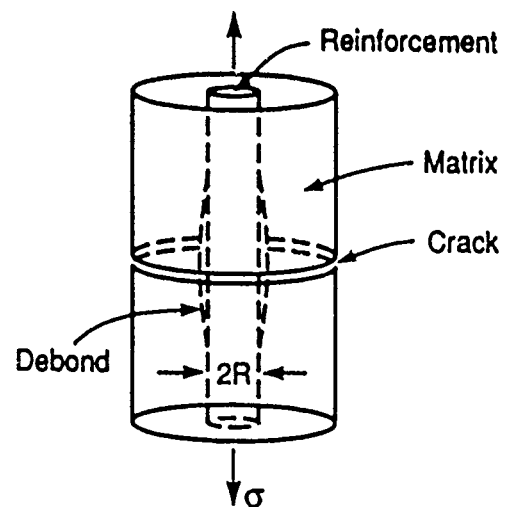
FIG. 7
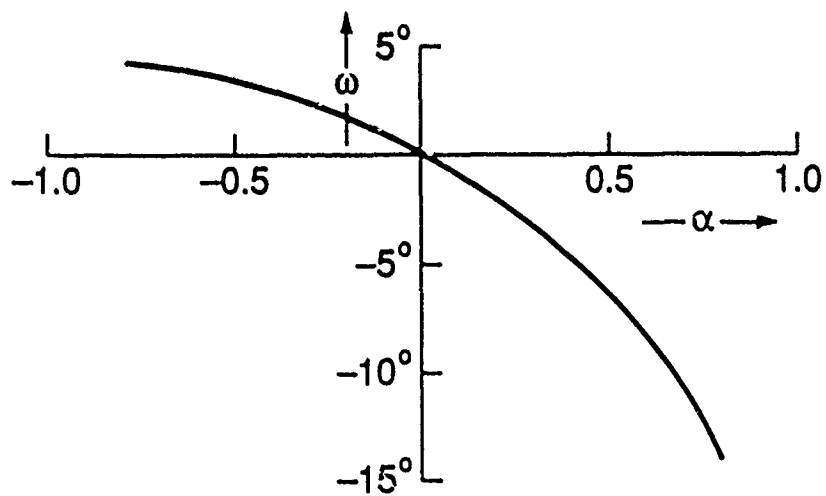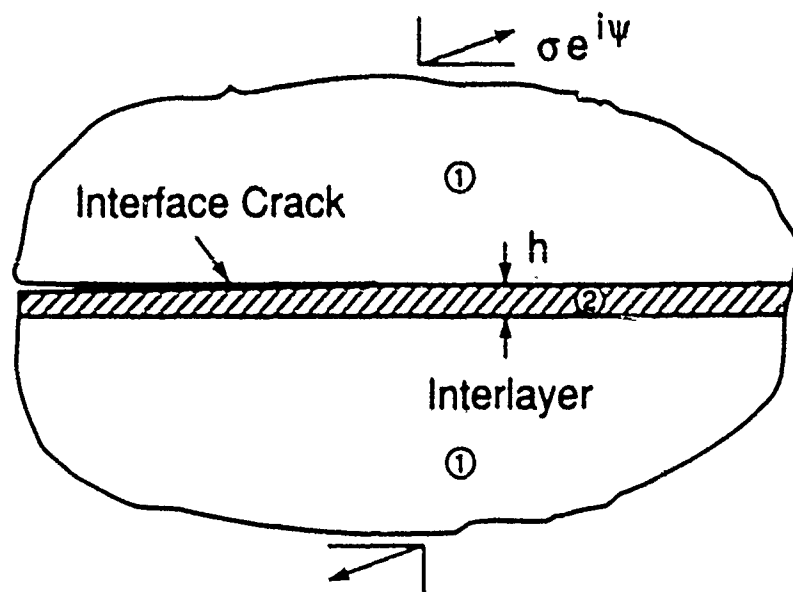
FIG. 8
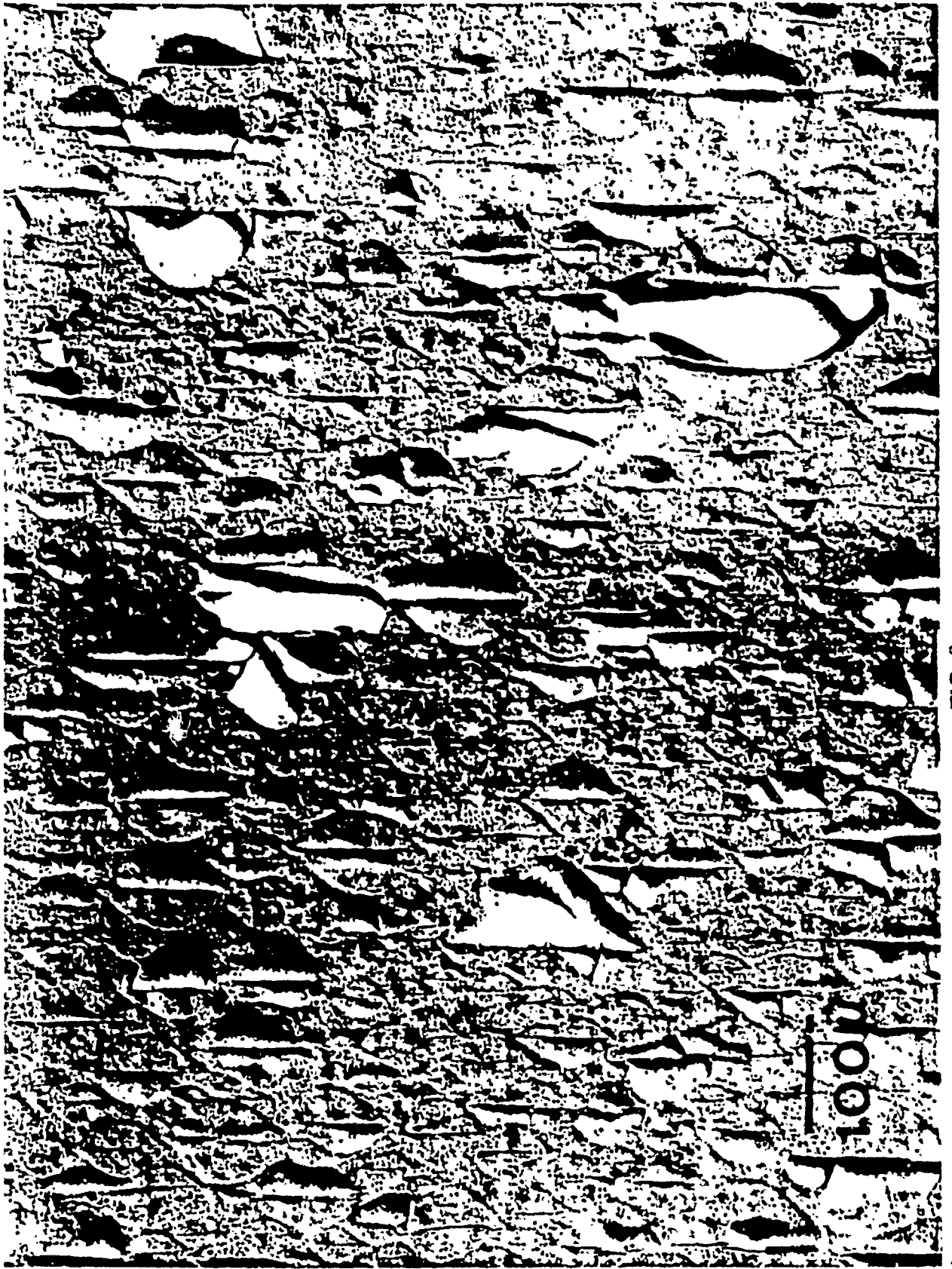
FIG. 9

100μ

MECHANICS AND THERMODYNAMICS OF BRITTLE
INTERFACIAL FAILURE IN BIMATERIAL SYSTEMS

by

James R. Rice, Zhigang Suo

and

Jian-Sheng Wang

Division of Applied Sciences
Harvard University
Cambridge, Massachusetts 02138

June 1989

MECHANICS AND THERMODYNAMICS OF BRITTLE
INTERFACIAL FAILURE IN BIMATERIAL SYSTEMS

James R. Rice, Zhigang Suo and Jian-Sheng Wang
Division of Applied Sciences
Harvard University, Cambridge, MA 02138 USA

June 1989

1. Introduction

We discuss elastic-brittle fracture theory for cracks along interfaces between elastically dissimilar solids. Crack tip fields in such cases are characterized by one real and one complex stress intensity factor, where the latter couples two of the classically separate crack tip modes. Solutions for complex stress intensity factors for a variety of cases, including geometries of interest for toughness testing, are cataloged and the theory is applied to the problem of dislocation nucleation from a crack tip along a dissimilar material interface, such as a metal/ceramic interface. The variation of the local phase angle, indicating stress field mode coupling, at the near atomic scale of dislocation nucleation is an important aspect of the problem. The dislocation problem arises in evaluating the competition between atomically brittle decohesion and plastic blunting at interfacial crack tips. Results confirm that both the properties of the interface and the direction of attempted cracking along it are important to the outcome of that competition; the latter arises because of the different stressing of slip systems associated with different directions of cracking. We also briefly review effects of solute adsorption at ceramic and metal/ceramic interfaces on their embrittlement, or in some cases on their ductilizing. Finally, we outline the thermodynamic inter-relations between adsorption and alterations of the ideal work of interfacial separation, and formulate the thermodynamic theory in a manner consistent with the possibility of strong dissimilarity in adsorption properties of the two surfaces which joined at the interface before fracture.

2 Elastic Fields for Interface Cracks

In this section we briefly outline the form of near-tip fields from linear elastic continuum solutions for cracks along an interface between dissimilar materials. Such cracks are most simply modeled as surfaces across which no tractions are transmitted. Solutions resulting for that model generally predict material interpenetration very near the tip. However, that feature can often be ignored in interpreting the solutions, either because the predicted contact zone is small compared to physical sizes of interest like atomic spacings (as is the case in later applications here to dislocation nucleation at a crack tip) or larger scale measures of material heterogeneity, or because the predicted contact region occurs well within a near tip region in which mechanical loading may be characterized in terms of the singular fields of such elastic solutions (Rice, 1988). The case of joined isotropic materials, most discussed here, was analyzed between 1959 and 1965 by Williams, Cherepanov, England, Erdogan, Rice and Sih, and Malyshev and

1

Salganik; it has been the subject of much recent work (for references see Rice, 1988, Hutchinson, 1989, and Suo, 1989a). The case of cracks between dissimilar anisotropic solids was first analyzed by Gotoh in 1967 and by Clements and Willis in 1971 and has been discussed recently by Anderson (1988), Bassani and Qu (1989), Suo (1989a,c) and Wu (1989). A list of stress intensity factors and available specimens will be presented at the end of this section, which is intended to serve as a self-contained mini-reference for practitioners of interfacial fracture mechanics.

2.1 Near tip stress fields. Ignoring terms which remain bounded at the tip, the near-tip stress field for an interface crack between dissimilar isotropic materials has the singular form (Fig. 1)

$$\sigma_{\alpha\beta} = \frac{1}{\sqrt{2\pi r}} \left[Re\,(K\,r^{i\varepsilon})\,\Sigma_{\alpha\beta}^{I}(\theta) + Im\,(K\,r^{i\varepsilon})\,\Sigma_{\alpha\beta}^{II}(\theta) + K_{III}\,\Sigma_{\alpha\beta}^{III}(\theta) \right] \quad (\alpha,\,\beta = x,\,y,\,z) \quad (2.1)$$

Here the angular functions $\Sigma_{\alpha\beta}(\theta)$ of superscripts I, II and III correspond to tractions across the interface at $\theta = 0$ of tensile, in-plane shear and anti-plane shear type, respectively, so that

$$(\sigma_{yy} + i\,\sigma_{yx})_{\theta=0} = \frac{K\,r^{i\varepsilon}}{\sqrt{2\pi r}}, \qquad (\sigma_{yz})_{\theta=0} = \frac{K_{III}}{\sqrt{2\pi r}}, \qquad (2.2)$$

and in this sense those $\Sigma_{\alpha\beta}(\theta)$ may be said to correspond to modes I, II and III. However $\Sigma_{\alpha\beta}^{I}(\theta)$ and $\Sigma_{\alpha\beta}^{II}(\theta)$ also depend on elastic properties of the bimaterial combination, through the parameter ε to be discussed, and do not have the full symmetry and anti-symmetry respectively associated with such angular functions for the case of a crack tip in a homogeneous solid. The functions are given in the Appendix and plotted in Fig.2. The parameter ε is given by

$$\varepsilon = \frac{1}{2\pi}\,ln\left[\frac{(3 - 4\nu_1)/\mu_1 + 1/\mu_2}{1/\mu_1 + (3 - 4\nu_2)/\mu_2}\right] \qquad (2.3)$$

where μ is shear modulus and ν the Poisson ratio, and subscripts refer to the two materials (Fig. 1).

The K_{III} is a mode III stress intensity factor of familiar type, but rather than having mode I and II factors, there is a single complex stress intensity factor K for the in-plane modes, and these mo.. ˙ ˙ e inherently coupled together. This complex K has the generic form

$$K = Y\,T\,\sqrt{L}\,L^{-i\varepsilon}\,e^{i\psi} \qquad (2.4)$$

(examples are given subsequently) where T is a representative magnitude of the stress applied to load the specimen, L a characteristic length (e. g., crack length, layer thickness), Y a dimensionless real positive quantity, and ψ by definition is the phase angle of $K\,L^{i\varepsilon}$, but is often loosely called the phase angle of the complex stress intensity factor, or the phase angle of the applied load. (A better characterization of the phase is given below.) Both Y and ψ are dependent on the detailed manner of applying the stress T and, in general, on the ratios of elastic moduli and of characteristic dimensions of the cracked body to one another.

Thus, considering eqs.(2.1,2), the analogs of K_I and K_{II} for the bimaterial case are not constants but are instead functions of r, that we might denote as $K_I(r)$ and $K_{II}(r)$, given by

$$K_I(r) = Re\,(K\,r^{i\varepsilon}) = Y\,T\,\sqrt{L}\,cos\,[\psi - \varepsilon\,ln\,(L/r)],$$

2

$$K_{II}(r) = Im\,(\,K\,r^{i\epsilon}\,) = Y\,T\,\sqrt{L}\,sin\,[\psi - \epsilon\,ln\,(L/r)]. \qquad (2.5)$$

Evidently the ratio $\sigma_{yx}/\sigma_{yy} = K_{II}(r)/K_{I}(r) = tan\,[\psi - \epsilon\,ln\,(L/r)]$ varies with r near the tip, so that the tensile and in-plane shear modes are fundamentally inseparable. The variable quantity $\psi - \epsilon\,ln\,(L/r)$ gives what might be called the *local phase angle* of the field; a similar concept was introduced by Liechti and Hanson (1988).

To better characterize phase, we may choose some fixed length $\hat{r}$. It does not matter whether we choose $\hat{r}$ as 1 nm, or 1 mm, or 1 m, or whatever, so long as we regard it as held fixed when specimens of a given bimaterial combination corresponding to a range of values of L and ψ are considered. Then a suitable absolute characterization, say $\hat{\psi}$, of the phase angle is given by writing

$$K_{I}(\hat{r}) + i\,K_{II}(\hat{r}) = K\,\hat{r}^{i\epsilon} = \sqrt{[K_{I}(\hat{r})]^2 + [K_{II}(\hat{r})]^2}\;e^{i\hat{\psi}} = Y\,T\,\sqrt{L}\,(\hat{r}/L)^{i\epsilon}e^{i\psi} \qquad (2.6)$$

so that

$$\hat{\psi} = \psi - \epsilon\,ln(\,L/\hat{r}\,). \qquad (2.7)$$

In citing the value of $\hat{\psi}$ to be associated with a given specimen, clarity requires that one also report the value of $\hat{r}$. A set of $\hat{\psi}$ values based on the fixed length $\hat{r}_A$ differs from a set based on another fixed length, $\hat{r}_B$, by a constant angle, i. e., $\hat{\psi}_A - \hat{\psi}_B = \epsilon\,ln(\hat{r}_A/\hat{r}_B)$, for every member of the set. If $\hat{r}$ is chosen as some length on the laboratory scale in crack studies, say $\hat{r} = 1$ mm, then for typical small ϵ (see Table I) the difference between $\hat{\psi}$ and ψ will be small for all L in a fairly broad range, say 0.1 to 10 mm.

In the above discussion there is no fundamental meaning attached to the quantities denoted by $K_{I}(\hat{r})$ and $K_{II}(\hat{r})$, since $\hat{r}$ is chosen arbitrarily. Nevertheless, because of the weak $ln(L/r)$ variation with r in eqs.(2.5), and because ϵ is typically very small, the trigometric functions of (2.5) are often little changed by fractionally large changes in r. E. g., for $\epsilon = 0.05$, which is fairly large, a factor of 5 change in r alters the arguments of the trigometric functions by $\epsilon\,ln\,5 \cong 0.08 \cong 5°$. Because of that, it can sometimes be justified to evaluate $K_{I}(r)$ and $K_{II}(r)$ at a fixed material distance $\hat{r}$, thought to be broadly representative of the size of the "fracture process zone" for a given bimaterial, and then to consider $K_{I}(\hat{r})$ and $K_{II}(\hat{r})$ as classical mode I and II intensity factors. Our subsequent treatment of dislocation nucleation at an interfacial crack tip can be interpreted in that way, based on an atomic sized $\hat{r}$. Of course, K rigorously reduces to the classical $K_{I} + i\,K_{II}$ only when $\epsilon = 0$, i. e., when β as defined below is zero.

Joined anisotropic solids. The singular stress field near the tip of an interface crack between dissimilar anisotropic solids can also be put in the form of eq.(2.1), in the sense that the field can be characterized by one real stress intensity factor and one complex factor, the latter multiplying a term $r^{i\epsilon}$ (e. g., Suo, 1989a,c, Wu, 1989), except that now the analogs of the angular functions Σ^{I}, Σ^{II}, and Σ^{III} of (2.1) no longer correspond, in general, to respective tractions of tensile, in-plane shear, and anti-plane shear type along the interface ahead of the tip.

When the material symmetries of joined anisotropic solids are such that in-plane loadings produce no

3

anti-plane deformations, and conversely, the real intensity factor K_{III} still corresponds to stresses of type σ_{yz} on the interface (i. e., to conventional mode III), as remarked by Bassani and Qu (1989). However, cases exist for which the real intensity corresponds to stress of type σ_{yy} ("mode I") on the interface, and for which σ_{yx} ("mode II") and σ_{yz} ("mode III") there are characterized by a complex K with $\epsilon \neq 0$, and hence are coupled and oscillatory when that $K \neq 0$. This situation occurs, in general, for interface cracks in symmetric tilt bicrystals (i. e., two crystals of the same material in mirror-image orientation relative to one another). By symmetry, the loading of such bicrystals solely by symmetric tensile forces relative to the interface produces only σ_{yy} and no σ_{yx} or σ_{yz} on the interface; hence such loadings produce complex $K = 0$ in that case and generate only a classical, non-oscillatory mode I singularity characterized by a real intensity factor. When the crack tip coincides with the tilt axis of such a symmetric tilt bicrystal, and when that common axis is also an axis of crystal symmetry such that in-plane and anti-plane deformations decouple (e. g., [100], [110], or [111] axes in cubic crystals), it turns out, as could be anticipated from the special cases discussed so far, that $\epsilon = 0$ and the entire crack tip stress state is then non-oscillatory under general loading (Bassani and Qu, 1989).

Ting (1986) has given the general formula for calculating ϵ in terms of the anisotropic modulus tensors of the two joined solids and also proven a remarkable theorem which shows that for a given pair of orientations of two dissimilar anisotropic crystals, ϵ is invariant to rotations of the interface plane about the crack tip. Since $\epsilon = 0$ in the special cases involving symmetric tilt bicrystals noted in the last sentence of the previous paragraph, it then follows from the Ting theorem that $\epsilon = 0$ also for interface cracks in bicrystals that are asymmetrically tilted about an axis of crystal symmetry for which in-plane and anti-plane deformations decouple, so long as the crack tip is along that tilt axis. In cases of interfacial cracks between anisotropic solids with $\epsilon = 0$, it need not follow in general that the complex K reduces to the classical $K_I + i\,K_{II}$ (Suo, 1989a.c).

2.2 **Values of ε for joined isotropic solids.** As observed by Dundurs (1969), for a bimaterial composite of isotropic constituents under plane strain conditions, with traction prescribed on the boundary, stresses in the body depend on only *two* dimensionless moduli combinations, α and β, defined as

$$\alpha = \frac{(1-v_2)/\mu_2 - (1-v_1)/\mu_1}{(1-v_2)/\mu_2 + (1-v_1)/\mu_1} \;, \quad \beta = \frac{1}{2}\frac{(1-2v_2)/\mu_2 - (1-2v_1)/\mu_1}{(1-v_2)/\mu_2 + (1-v_1)/\mu_1} \quad \left[\epsilon = \frac{1}{2\pi} ln\!\left(\frac{1-\beta}{1+\beta}\right) \right]; \qquad (2.8)$$

α measures the relative stiffness of the two materials. By requiring $0 < v < 1/2$ and $\mu > 0$, Dundurs showed that α and β are confined to a parallelogram in the (α, β) plane with vertices at $(1, 0)$, $(1, 0.5)$, $(-1, 0)$ and $(-1, -0.5)$, so that the maximum possible $|\epsilon| = 0.175$. Calculations by Suga et al. (1988) on more than a hundred material pairs suggest that the values of β are even more restricted, i.e., $|\beta| < .25$, implying that $|\epsilon| < 0.08$. Representative values of these parameters are given in Table I. The elastic constants used for the calculations are taken from Hirth and Lothe (1982) for all single elements and Suga et. al (1988) for sapphire (Al_2O_3) fiber and MgO. Note that α, β and ϵ reverse sign when materials 1 and 2 are interchanged.

4

TABLE I   Dundurs parameters and the oscillatory index ϵ.

bimaterial (mat'l 1 / mat'l 2)	α	β	ϵ
Al /Sapphire	-.69	-.143	.046
Au /Sapphire	-.62	-.053	.017
Cu /Sapphire	-.47	-.096	.031
Nb /Sapphire	-.55	-.056	.018
Ni /Sapphire	-.26	-.063	.020
Fe /Sapphire	-.30	-.065	.021
Au /MgO	-.53	-.062	.020
Ni /MgO	-.13	-.079	.025
Cu /Si	-.04	.038	-.012

2.3 **Catalog of solutions for complex K.** *i) An internal crack:* Analyses of collinear interface cracks between two half-planes of dissimilar materials were the subject of 1965 papers by England, Erdogan, and Rice and Sih. As an example, for an internal crack of length 2a with stress state $(\sigma_{yy}^{\infty}, \sigma_{yx}^{\infty})$ at infinity (Fig. 3), the stress intensity factor at the *right-hand* crack tip is

$$K = (1 + 2i\ \epsilon)\left(\sigma_{yy}^{\infty} + i\ \sigma_{yx}^{\infty}\right)\sqrt{\pi a}\ (2a)^{-i\epsilon}$$

$$= T e^{i\ [\omega + arctan\ (2\epsilon)]}\ \sqrt{(1 + 4\epsilon^2)\ \pi a}\ (2a)^{-i\epsilon} \qquad (2.9)$$

We have used $\sigma_{yy}^{\infty} + i\ \sigma_{yx}^{\infty} = T e^{i\omega}$ in the latter form, where T is the magnitude of the remotely applied traction vector and ω is its angle with the y direction, to give agreement with the form of eq.(2.4). Thus the phase angle ψ in this case is $\omega + arctan(2\epsilon) \cong \omega$. Notice that the material dependence of K in (2.7) is through ϵ only. This is a feature common to all problems of collinear interface cracks between two half planes under self-equilibrated traction on crack faces.

ii) Crack/dislocation interactions: The problems of cracks interacting with various singularities (dislocation, line force, transformation spot, etc.) are solved in Suo (1989a,b). The stress intensity factor, for instance, of a semi-infinite crack due to a dislocation with Burgers vector $b = b_y + i\ b_x$ at point $s = x_0 + i\ y_0$ in material 2 (Fig. 4) is

$$K = -\frac{E^* \cosh \pi\epsilon}{4\sqrt{2\pi}}\left[\left(e^{-\pi\epsilon}\ \bar{s}^{-1/2-i\epsilon} + e^{\pi\epsilon}\ s^{-1/2-i\epsilon}\right)b + (1/2 + i\epsilon)\ e^{\pi\epsilon}\ (s - \bar{s})\ s^{-3/2-i\epsilon}\ \bar{b}\right] \qquad (2.10)$$

where $1/E^*$ is the average of $(1-\nu)/2\mu$ for the two materials. One must change $\pi\epsilon$ to $-\pi\epsilon$ everywhere to obtain K when the dislocation is in material 1.

iii) A thin film under residual tension: Consider a thin film of thickness h, deposited on an elastic substrate, with residual tensile stress σ in the film (Fig. 5a). The induced stress intensity factor for a long crack is given by

$$K = \sqrt{\frac{1 - \alpha}{2(1 - \beta^2)}}\ \sigma\ h^{1/2 - i\epsilon}\ e^{i\psi(\alpha, \beta)} \qquad (2.11)$$

5

where α and β are the Dundurs' parameters. The phase angle $\psi(\alpha, \beta)$ is plotted in Fig. 5b for $\beta = 0$ (the error due to $\beta \neq 0$ is small). This problem was solved in Suo and Hutchinson (1989a). The thickness of the substrate is assumed to be infinite in (2.11). Solutions for finite substrates can be found in the original paper. Argon et al. (1989) have used residual stresses in thin films as driving forces to measure interface toughness.

iv) Sandwich specimens: Any homogeneous specimen can be converted to measure interface toughness by sandwiching a *thin* second material layer of thickness h, Fig.6a. The residual stress in the thin layer does not contribute to K and, in calibrating such a specimen, one needs to take account of the external loading only. This arrangement has been experimentally tested by Oh et al. (1987) and Cao and Evans (1989) with various reference homogeneous specimens. The measured critical loads are first used to calculate the classical stress intensity factors K_I and K_{II} as if the specimen were homogeneous (i. e., as if $h = 0$) using, for instance, the standard handbook solutions. The latter are then converted to the interface stress intensity factor K. A universal conversion relation, which is independent of the geometry of the reference homogeneous specimen when h is small, is (Suo and Hutchinson, 1989b)

$$K = \sqrt{\frac{1-\alpha}{1-\beta^2}} \left(K_I + i\, K_{II} \right) h^{-i\varepsilon}\, e^{\,i\,\omega\,(\alpha,\,\beta)} \tag{2.12}$$

The phase angle shift relative to the classical stress intensity factor, $\omega(\alpha, \beta)$, is typically small and is 1 in Fig. 6b with $\beta = 0$.

) A bending specimen: The specimen illustrated in Fig.7a was analyzed by Charalambides et al. (1989a,b) and Suo and Hutchinson (1989a). When the crack is long compared with the thickness of the notched layer, h, the stress intensity factor is of the form

$$K = Y(\alpha, \beta, h/H)\, M\, h^{-3/2-i\varepsilon}\, e^{\,i\,\psi\,(\alpha,\,\beta,\,h/H)} \tag{2.13}$$

where M is the moment per unit width, and Y and ψ as the functions of α and h/H are plotted in Figs. 7b and 7c, with $\beta = 0$. Thermal mismatch stress between the two layers is a driving force for crack extension and analysis of that can be found in the above cited papers.

2.4 Crack opening displacement and energy release rate. Letting δ_α denote components of the relative displacement of two initially coincident points along the crack walls, for isotropic solids

$$\delta_y + i\, \delta_x = \left(\frac{1-\nu_1}{\mu_1} + \frac{1-\nu_2}{\mu_2}\right) \frac{K\, r^{i\varepsilon}\sqrt{2r}}{\sqrt{\pi}\,\cosh(\pi\varepsilon)\,(1+2i\,\varepsilon)}\ , \quad \delta_z = \left(\frac{1}{\mu_1} + \frac{1}{\mu_2}\right) \frac{K_{III}\sqrt{2r}}{\sqrt{\pi}} \tag{2.14}$$

very near the tip, and the Irwin-type energy release rate expression is

$$G = \left(\frac{1-\nu_1}{\mu_1} + \frac{1-\nu_2}{\mu_2}\right) \frac{K\,\overline{K}}{4\cosh^2(\pi\varepsilon)} + \left(\frac{1}{\mu_1} + \frac{1}{\mu_2}\right) \frac{K_{III}^2}{4} \tag{2.15}$$

where we note that $K\,\overline{K} = K r^{i\varepsilon}\,\overline{K r^{i\varepsilon}} = [K_I(r)]^2 + [K_{II}(r)]^2 = Y^2\, T^2\, L$ in the notation of eq.(2.4).

2.5 Contact and range of validity of complex K. Since eqs.(2.5) show that $K r^{i\varepsilon}$ oscillates with r as $r \to 0$, (2.14) shows that the condition $\delta_y < 0$ (i. e., predicted interpenetration) must occur for some

6

sufficiently small r. We can estimate the distance r_{con} over which contact occurs near the tip as approximately twice the r at which δ_y or σ_{yy} is first predicted to turn negative by the solutions that ignore contact (more precise analyses of the near tip contact problem, within linear elastic theory, have been given by Comninou, 1977, and Comninou and Schmueser, 1979). Thus for small $\varepsilon > 0$

$$r_{con} \approx 2L \, exp\{-(\psi + \pi/2)/\varepsilon\} \,, \tag{2.16}$$

and one just changes ε to $-\varepsilon$, and ψ to $-\psi$ in this equation when $\varepsilon < 0$.

The predicted r_{con} is sometimes of subatomic size. For example, tak $L = 5mm$, we may ask how small ε must be for the predicted r_{con} to be less than an atom spacing ($\approx$ 5 nm). Considering material combinations with $\varepsilon > 0$, the condition is met whenever $\varepsilon < 0.069$ if $\psi \sim +45°$, whenever $\varepsilon < 0.046$ if $\psi = 0°$, and whenever $\varepsilon < 0.023$ if $\psi = -45°$. These conditions are met by most of the combinations in Table I, and by all if we consider only angles $\psi > 0$. When $\varepsilon > 0$, r_{con} becomes negligible (subatomic, for any laboratory sized L) as $\psi \to 90°$, but r_{con} becomes of order L as $\psi \to -90°$. The converse holds for combinations with $\varepsilon < 0$.

The predicted near tip contact, of which the scale is not always subatomic as above, has generally been regarded as an obstacle to the development of interface fracture mechanics. Clearly, the forms given here for the near tip stress and displacement fields must be wrong on some sufficiently small scale. However, as Rice (1988) has recently emphasized in discussing the plane strain state, the parameter K entering those fields does nevertheless characterize the severity of the near-tip loading whenever the size of a zone involving any or all of interface contact, nonlinear material response (plasticity, smaller scale microcracking, transformation), or dominance by discreteness of material microstructure (atom spacing in a crystal, fiber spacing in a composite treated globally as a continuum, etc.) is much smaller than characteristic macroscale dimensions like crack length. Thus just as K_I, K_{II} and K_{III}, and their histories, uniquely characterize the near tip state in conventional fracture mechanics, whenever such a size restriction is met, so also do (complex) K and K_{III} in interface fracture mechanics. In fact, the singular stress and displacement fields discussed here provide an intermediate asymptotic description of those fields over a range of r much smaller than a characteristic length like L but much larger than the near tip zone dominated by contact, nonlinearity and/or discreteness; the K concept is valid when such an intermediate length scale exists.

2.6 Fracture toughness for interface cracks. Consider a bimaterial specimen with a pre-existing interface crack and assume for simplicity that it is loaded so as to induce only in-plane stresses of type σ_{yy} and σ_{yx} (but no mode III stress σ_{yz}) on the interface ahead of the tip. Since tension and shear are inherently coupled for the interface crack, there can be no unambiguous characterization of a mode I and mode II toughness. Instead the toughness must be represented as a magnitude $|K|$ ($= \sqrt{K \bar{K}} = Y T \sqrt{L}$) of K for each value of the phase angle of interest, and the phase angle is properly reported as $\hat{\psi}$ of (2.6) based on a chosen (and specified) $\hat{r}$. One loads the specimen proportionally (thus ψ and $\hat{\psi}$ are fixed), records the magnitude of the loading stress, T, at the onset of crack propagation, and then, with a relation

7

of form (2.4) appropriate for the specimen, calculates the corresponding |K|. Repeating this procedure systematically for various $\hat{\psi}$, e.g., by changing the relative proportions of various loads, or by changing crack length or specimen design, one obtains a $|K|_c$-$\hat{\psi}$ curve, referred to as the *toughness locus* of the interface, where the subscript c denotes the critical value. Within the framework of elastic fracture mechanics, this curve is a property of the interface, which is independent of the specimen geometry and loading system. Since |K| is related to the energy release rate G by (2.15), it is equivalent to report a G_c-$\hat{\psi}$ curve instead.

Clearly, the oscillating singularity of the interface crack problem does not present any essential difficulty in engineering application of the macroscopic fracture mechanics methodology. Rather, it is the *mode mixity* that complicates the method. A crack in an isotropic, homogeneous brittle solid can usually adjust itself onto a path with the mode I state at the tip. For this reason, classical fracture mechanics requires information primarily about only one material property, namely, K_{Ic}. By contrast, since an interface is often an easy fracture path, compared to one through the adjoining bulk materials, a crack, even if subjected to substantial shear stresses ahead the tip, may tend to stay along the interface. Consequently, instead of one material property, one needs a *curve*, or toughness locus, to fully characterize the toughness of an interface.

Experimental determination of interface toughness has been attempted recently on several model systems (Argon et al., 1989, Cao and Evans, 1989, Charalambides et al., 1989a,b, Oh et al., 1987), with only one of them (Cao and Evans) covering a relatively wide range of mode mixity. In principle, any geometry with an interfacial crack, with a calibration relation in the form of (2.4) connecting measurable driving forces and stress intensity factors, may be used as a specimen for measuring interface toughness. Thin films under residual tension, sandwiches, and bending specimens, all discussed in section 2.3, seem to be especially suitable for measuring toughness of brittle interfaces.

3. Crack Tip Competition: Interface Cleavage versus Dislocation Emission

The Kelly-Tyson-Cottrell (Kelly et al, 1967) and Rice-Thomson approaches (Rice and Thomson, 1974) have focussed on the competition of dislocation emission and atomic decohesion at a crack tip as a test of whether a given ductile crystal is intrinsically cleavable. The R-T approach has also been extended to interfaces between ductile crystals (Rice, 1976; Mason, 1979; Anderson and Rice, 1986; Rice, 1987; Anderson et al., 1989; Wang and Anderson, 1989) but, so far, no account has been included of effects like those discussed in the last section of dissimilarity of elastic properties across such an interface. We do so here in the form of a simple version of the R-T model, and apply it to certain orientations of ductile fcc metal crystals that are joined to brittle ceramics.

In recent evaluations of the R-T model starting with Mason (1979), the energy release rate calculated as necessary for dislocation emission from a crack tip, G_{disl}, is compared to that for cleavage, G_{cleav}, where the latter is understood as atomistically brittle decohesion of the interface. If $G_{cleav} > G_{disl}$, one expects that a dislocation will emanate from the crack tip, producing a blunted crack and spoiling the stress

8

concentration, before conditions for atomic decohesion can occur. It is then assumed that failure, if to occur from processes at that crack tip, must be by some ductile mechanism, at least when there is adequate mobility of dislocations in moving away from the crack tip once they have been generated there. Conversely, if $G_{disl} > G_{cleav}$, the crack tip remains stable against dislocation blunting, i. e., atomically sharp, and it is assumed that atomic decohesion can occur leading to a cleavage-like brittle interfacial fracture. We then say that the interface is intrinsically cleavable from that crack tip. The qualifyer "intrinsically" emphasizes that in materials capable of plastic flow the matter of whether cleavage will actually occur is strongly dependent on plastic flow properties (and hence on hardness, loading rate and temperature). This is because the ease of plastic flow in material surrounding the crack tip which will control whether that tip remains atomically sharp against externally generated dislocations that may move towards it along slip planes, and also whether the surroundings can support enough stress to actually allow the condition $G = G_{cleav}$ to be attained locally at such an atomically sharp crack. Metal/ceramic interfaces that are judged not to be intrinsically cleavable may, of course, sometimes fail by atomically brittle cracking near the interface in the ceramic.

Aspects of interfacial fracture other than those reflecting dissimilarity of properties of the two adjoining materials, which is the focus here, have been reviewed recently by Rice and Wang (1989) and the reader is referred there for further discussion.

3.1 Cleavage decohesion. The simplest model of cleavage decohesion is provided by the elastic-brittle Griffith model, which estimates G_{cleav} as the work $2\gamma_{int}$ of reversibly separating unit area of interface:

$$G_{cleav} = 2\gamma_{int} \tag{3.1}$$

where, in absence of matter diffusion to or from the interface during separation (see section 4),

$$2\gamma_{int} = f_{s,1} + f_{s,2} - f_{1/2} \tag{3.2}$$

Here subscripts 1 and 2 refer to the two solids, $f_{s,1}$ and $f_{s,2}$ are surface free energies for the faces of solids 1 and 2 that are exposed by fracture and $f_{1/2}$ is the free energy of the 1/2 interface existing before fracture.

Cohesive zone models incorporate some further details of the separation process. In the simplest versions, containing no account of lattice discreteness in directions parallel to the interface, the interfacial region is represented as two joined elastic continua which interact with one another such that a stress versus separation relation, $\sigma_{y\alpha} = \partial\phi(\delta_x, \delta_y, \delta_z)/\partial\delta_\alpha$ ($\alpha = x, y, z$) applies along the gradually decohering interface. Here ϕ is the potential of the cohesion, $\phi(0, 0, 0)$ corresponds to $f_{1/2}$, and $\phi(\delta_x, \infty, \delta_z)$ to $f_{s,1} + f_{s,2}$. Provided that the two crack walls can be pulled apart quasistatically to full separation, in a crack-like mode, without local dynamic instabilities or dislocation or other inelastic processes within the two solids, a well known argument based on the J integral shows that this model gives the same result as (3.1,2), at least in typical circumstances when the zone undergoing decohesion is much shorter than the overall crack length (e. g., Rice, 1987).

However, effects relating to lattice discreteness cause local instabilities during separation and make $2\gamma_{int}$ a lower bound to G_{cleav} even when no dislocations are left in the material after passage of the crack.

9

One such effect is "lattice trapping"; current estimates (e. g., Thomson, 1983, 1986) suggest that it only slightly raises G_{cleav} above $2\gamma_{int}$. Another effect, which has not yet been quantified but may be important given the inevitably coupled x and y direction displacing that will occur during interfacial separation between materials of strong dissimilarity in elastic properties (in which effects relating to the presence of ϵ of the last section are non-negligible), involves local instabilities in shear displacement effectively corresponding to dislocation formation in the interface region as it is being pulled apart.

3.2 <u>Dislocation emission</u>. Conditions for dislocation emission on a partic ilar slip system are not fundamentally expressible as a critical value of G, but rather as some critical value of a linear combination of K_I, K_{II}, and K_{III} (or their generalizations like in eqs.2.5) which characterize the shear stress along that slip plane. Thus while G_{cleav} may be regarded as constant, at least within the Griffith model, G_{disl} will not have a fixed value but, rather, will depend on the ratios of the K's to one another.

For example, the shear stresses on a slip plane which intersects the interface along the crack tip (fig. 8) are $\sigma_{\theta r}$ and $\sigma_{\theta z}$. Resolving these in the direction of the Burgers vector b for that system, with edge and screw components v_r and b_z, respectively, gives the force driving a straight dislocation line away from the crack tip as

$$f_{emit} = \sigma_{\theta r} b_r + \sigma_{\theta z} b_z = \frac{b}{r^{1/2}} \left[Re (K r^{i\epsilon}) S_I + Im (K r^{i\epsilon}) S_{II} + K_{III} S_{III} \right] \tag{3.3}$$

where b is the magnitude of b, the dislocation line is at distance r from the tip, the $\sigma_{\alpha\beta}$ (r, β = r, θ, z) are the stresses of (2.1), and each S_J is defined by

$$S_J = [(b_r/b) \Sigma_{\theta r}^J(\theta) + (b_z/b) \Sigma_{\theta z}^J(\theta)] / \sqrt{2\pi} \quad (J = I, II, III) . \tag{3.4}$$

Here θ is the angle of the slip plane (fig. 8) and the $\Sigma_{\alpha\beta}$ are from (2.1). For joined isotropic solids

$$S_I = \Sigma_{r\theta}^I(\theta) \cos \phi / \sqrt{2\pi} , \quad S_{II} = \Sigma_{r\theta}^{II}(\theta) \cos \phi / \sqrt{2\pi} , \quad S_{III} = \cos \theta/2 \sin \phi / \sqrt{2\pi} \tag{3.5}$$

where $\Sigma_{r\theta}^I(\theta)$ and $\Sigma_{r\theta}^{II}(\theta)$ are given in the Appendix and plotted in fig. 2, and ϕ is the angle which the Burgers vector makes with the normal to the crack tip, fig.8. Eqs.(3.3,4) are also valid for joined anisotropic solids with the interpretations given earlier.

The "image" force component opposing outward motion of the dislocation line, and drawing it back into the tip, has been solved for by Rice (1985) for the general case of joined anisotropic solids. Remarkably, it depends only on the elastic properties of the material (1, in the present case) in which the dislocation resides, and is

$$f_{image} = \frac{1}{r} b_\alpha A_{\alpha\beta} b_\beta \tag{3.6}$$

where $A_{\alpha\beta}$ is the pre-logarithmic energy factor matrix of a straight dislocation line in that material, i. e., $b_\alpha A_{\alpha\beta} b_\beta$ is the factor which appears when the energy per unit length of dislocation line is written as $E_{dislocation} = b_\alpha A_{\alpha\beta} b_\beta \ln (r_{outer}/r_{core})$. When material 1 is isotropic

$$A_{rr} = A_{\theta\theta} = \frac{\mu_1}{4\pi (1 - \nu_1)} , \quad A_{zz} = \frac{\mu_1}{4\pi} , \quad \text{other } A_{\alpha\beta} = 0 . \tag{3.7}$$

Thus, using the simplified emission criterion (e. g., Thomson, 1986) that $f_{emit} = f_{image}$ for $r = r_c$, the dislocation core size, which also assures that $f_{emit} > f_{image}$ for all $r > r_c$, the critical combination of intensity factors estimated as necessary for dislocation emission is given by

$$\left(Re \, (K \, b^{i\epsilon}) \, S_I + Im \, (K \, b^{i\epsilon}) \, S_{II} + K_{III} \, S_{III} \right)_{disl} = b_\alpha A_{\alpha\beta} b_\beta / [\, b \, r_c^{1/2} \,]$$

$$\left(= \mu_1 \, b \, [cos^2\phi + (1-v_1) \, sin^2\phi \,] / [\, 4\pi \, (1-v_1) \, r_c^{1/2} \,] \quad \text{for isotropic solid 1} \right) \qquad (3.8)$$

Here we have taken advantage of the fact that $K \, r^{i\epsilon}$ is a slowly varying function of r, given the small values of ϵ, in the sense that $K \, r^{i\epsilon} \cong K \, b^{i\epsilon}$ for all r of order b and, since core size r_c will always be of order b, we have just replaced $K \, r^{i\epsilon}$ by $K \, b^{i\epsilon}$ in (3.3).

Atomic scale phase angle ψ'. We may write

$$K \, b^{i\epsilon} = |K| \, e^{i\psi'} = cosh \, \pi\epsilon \, \sqrt{G \, E^*} \, e^{i\psi'} \qquad (3.9)$$

where, again, $1/E^*$ is the average of $(1-v)/2\mu$, so that this equation defines an *atomic scale phase angle*, denoted as ψ'. It is related to the angle ψ in the generic form of (2.4) for K by

$$\psi' = \psi - \epsilon \, ln \, (L/b) \, . \qquad (3.10)$$

This ψ' is just a special version of $\hat{\psi}$ of eq.(2.7) but, unlike the typical situation for $\hat{\psi}$ when the $\hat{r}$ on which it is based is chosen to be on the laboratory scale, the difference between ψ' and ψ will typically be significant. This is because L is generally macroscopic and b is of order 0.25 nm. For example, if $L \cong 5$ mm, $\psi' \cong \psi - 17 \, \epsilon$, which is a significant change from ψ even for the smaller ϵ in Table I. As a consistency check, one should assure that $-\pi/2 < \psi' < +\pi/2$ for use of the above expressions for dislocation nucleation; the condition effectively assures that interpenetration is not predicted over a scale as large as a lattice spacing near the crack tip.

Finally, considering joined isotropic solids under in-plane loadings, so that $K_{III} = 0$ and the near tip field is fully characterized by K, G_{disl} may be calculated from (3.8) and (2.15). After using (2.8) to rearrange some terms, the result is

$$G_{disl} = \frac{\mu_1 \, b^2}{(1 - v_1)(1 - \alpha) \, r_c} \left[\frac{cos \, \phi + (1 - v_1) \, sin \, \phi \, tan \, \phi}{4 \, \sqrt{\pi} \, cosh \, \pi\epsilon \, \left(\Sigma_{r\theta}^I(\theta) \, cos \, \psi' + \Sigma_{r\theta}^{II}(\theta) \, sin \, \psi' \right)} \right]^2 \qquad (3.11)$$

It is seen that G_{disl} depends on the phase angle ψ', and hence on the mode mixity of the applied loadings. In fact, the above result applies only for the range of ψ' such that $\Sigma_{r\theta}^I(\theta) \, cos \, \psi' + \Sigma_{r\theta}^{II}(\theta) \, sin \, \psi' > 0$; otherwise, the direction of shearing along the slip plane considered does not tend to drive a dislocation away from the tip, and emission cannot occur on that plane. Since G_{cleav} is independent of ψ', at least for the Griffith model, it may occur that $G_{disl} < G_{cleav}$ for one range of ψ' and $G_{disl} > G_{cleav}$ for another. Thus the nature of crack tip response is not universal but depends on the mode mixity of the applied loading.

Previous studies based on non-oscillatory crack tip fields and semi-circular loop model. We note that the modeling of dislocation emission from an interfacial crack tip is somewhat more advanced in cases for which there is a conventional non-oscillatory singularity at the tip. These cases are equivalent to setting ϵ

11

≈ 0 in the representations of the near-tip field here and letting K denote the classical $K_I + i K_{II}$. Such cases include symmetric tilt bicrystals as discussed earlier.

Here we summarize recent studies modeling the emission of a semi-circular dislocation loop from a crack tip (Anderson, 1986; Anderson and Rice, 1986; Wang et al., 1987; Wang and Anderson, 1989). These studies proceed like in Mason (1979)- calculating the energy U of a semi-circular loop of radius R introduced into a stressed solid at the crack tip, and determine R, as well as the combination $K_I S_I + K_{II} S_{II} + K_{III} S_{III}$ (with the S_J defined as in eq. 3.4), at nucleation from the conditions that $\partial U/\partial R = 0$ and $\partial^2 U/\partial R^2 = 0$. The result is

$$(K_I S_I + K_{II} S_{II} + K_{III} S_{III})_{dist} = 0.76 \frac{A_o}{b} \sqrt{\frac{m}{r_c}} \, exp \, (E_{ledge}/ \pi A_o) \tag{3.12}$$

Here A_o is the pre-logarithmic energy factor appearing in the expression $U = 2\pi R A_o ln(8R/e^2 r_c)$ for the energy of a full circular shear loop of the same orientation in an uncracked solid; $A_o = (2-v)\mu b^2/8\pi(1-v)$ ≅ $\mu b^2/10$ for an isotropic solid. E_{ledge} is the energy of the ledge produced at the blunted crack front due to dislocation emission; Anderson (1986) estimates $E_{ledge}/\pi A_o$ ≈ 0.1 to 0.6 for various cubic metals. The factor m has been calculated for shear loops at a crack tip by Gao and Rice (1989); it arises in the exact (within continuum elasticity) expression $U = \pi R A_o ln(8mR/e^2 r_c)$ for the self energy of a semi-circular shear loop of radius R emanating from the crack tip; m depends on θ and ϕ (fig. 8) with typical values in the range from about 1 to 2.

For some materials dislocations may be nucleated in dissociated form as a pair of partials. In that case terms related to the stacking fault energy and the interaction energy between the partials must be included (Anderson, 1986; Anderson and Rice, 1986; Wang and Anderson, 1989). Taking m = 1.5, $E_{ledge}/\pi A_o$ = 0.3 and ϕ = 30° as representative, and using expressions for identical isotropic solids with v = 0.3, predictions from the simpler model leading to (3.8) and the presumably more exact expression of (3.12) can be compared. The results are fairly close to one another; the former gives $0.11\mu b/r_c^{1/2}$ and the latter $0.13\mu b/r_c^{1/2}$ for the value of $K_I S_I + K_{II} S_{II} + K_{III} S_{III}$ at nucleation.

Bicrystal experiments. Although as noted earlier, the R-T model does not address important facets of ductile versus brittle failure relating to viscoplastic material response, the model has enabled rationalization of observed intrinsic behavior for various systems. These include the orientation and cracking direction dependence for a series of [110] symmetric tilt Cu-Bi alloy bicrystals, with interfacial cracks introduced in directions such that their tips are along the tilt axis, as studied recently by Wang (1988, 1989). Predictions of G_{dist} and G_{cleav} for the bicrystals tested were made by Wang and Anderson (1989) based on a formulation like in eq.(3.12) and with structural and segregation models for $2\gamma_{int}$. Behavior of the Σ9 [110] symmetric tilt bicrystal, with (2$\bar{2}$1) interfacial plane, is especially interesting. It can be cracked under tensile loading in a brittle manner along the interface when the direction of crack introduction is [1$\bar{1}$4], but the interface does not crack and failure is instead ductile when the direction of attempted cracking is in the opposite direction, i. e., in the [$\bar{1}$14] direction. Theoretical calculations based on

expressions like (3.12) have been shown (Anderson and Rice, 1986; Wang and Anderson 1989) to predict a much higher value of G_{dist} (under the same pure mode I loading) for the $[1\bar{1}4]$ cracking direction than for the $[\bar{1}14]$ direction; this is due to the different orientations of potentially relaxing slip systems relative to the crack tip in the two cases. Accordingly, the observed behavior can be explained as a case of G_{cleav} being reached before G_{dist} for the brittle $[1\bar{1}4]$ direction, but G_{dist} before G_{cleav} for the ductile $[\bar{1}14]$ direction. Thus, as Rice and Wang (1989) pointed out, it oversimplifies to say that an interface is intrinsically cleavable, because G_{dist} is not a fixed number for a given interface but depends on the direction of crack growth along it, as well as on the mixity of shear with tensile loading relative to the crack. Similar orientation and cracking direction dependence is to be expected for interfaces between dissimilar materials, such as typical metal/ceramic interfaces, and these effect are embodied by our expression in (3.8) for G_{dist}. We illustrate them in the next subsection in analysis of Cu/Al$_2$O$_3$ interfacial fracture specimens.

3.3 Application to single crystal Cu/sapphire interfaces. Consider a four-point bending specimen of a Cu single crystal bonded to a layer of sapphire (fig. 9a,b). The sapphire layer is notch cracked to the interface and a central crack continues along the interface. The Cu single crystal is oriented in all cases that we consider so that the crack fronts at both ends lie along the intersection of a pair of (111) slip planes of the Cu crystal with the interface; i. e., the crack tip lies along [110] in the Cu in all cases. This configuration allows application of the simple model developed in the preceding section, summarized as eqs. (3.8, 10, 11).

For simplicity, and because some of the solutions needed would not otherwise be available, we neglect the elastic anisotropy of the Cu and Al$_2$O$_3$ here. Both are treated as elastically isotropic, and we use averaged properties for each as in preparing Table I. The thickness h of the sapphire is assumed to be 1 mm in fig. 9a and b, and the ratio of Cu to sapphire thickness to be 3. Also, the crack length is assumed long enough for the analysis summarized by eq. (2.13) and fig. 7 to apply. For the present case, $\alpha = -0.47$ and $\epsilon = 0.031$. From fig. 7c, $\psi = -52°$ at the two crack tips in all cases (we reverse the signs of α, ϵ and ψ in reading fig. 7 since the 1 and 2 labelling of the materials is interchanged there; here 1 is Cu and 2 is Al$_2$O$_3$). Thus the phase angle ψ' of (3.10) relevant to crack tip events at the atomic scale is, since L = h for the definition of ψ in (2.13),

$$\psi' = \psi - \epsilon \, ln \, (h/b) = -52° - 0.031 \, ln \, (1mm/0.25nm) \approx -52° -27° = -79° \qquad (3.13)$$

This is a significant phase shift. (The effects of non-zero ϵ, and of the associated variable loading phase conditions near the crack tip, can often be neglected in engineering applications of interface fracture mechanics, as suggested by Hutchinson (1989), especially in a small ϵ situation like the present case. However, since we are here dealing with a process occurring at the atomic scale, the predicted shift of the phase angle (-27°) relating to ϵ is significant and has some possibly testable consequences. The entire ϵ effect of consequence is, however, only in the shift of the phase angle; the small non-zero ϵ of 0.031 has negligible effect on the functions $\Sigma_{r\theta}(\theta)$, as can be seen from fig. 2.

Values of G_{dist} have been calculated from (3.11) for various crack tips in fig. 9, and results are

13

summarized below. The favored of the two (111) slip planes passing through each tip is that on which the combination involving ψ' in the denominator in (3.11) has the largest value; if ψ' is such that the combination is negative for both systems, then dislocations cannot be nucleated on them. We assumed that the dislocations nucleate in dissociated form and that the controlling event is the nucleation of the second partial. This is treated approximately by evaluating (3.11) with $\phi = 60°$ and $r_c = b = 0.15$ nm for the partial (and with $\mu_1 = 5 \times 10^{10}$ N/m^2 and $v_1 = 0.3$ for Cu). We note that the comparison of (3.8) with (3.12) above suggests that the actual G_{disl} may be of order 50% higher than what we estimate from (3.11); also, G_{disl} is proportional to b/r_c, and the proper choice of core size r_c is quite uncertain. Thus the magnitudes that we report for G_{disl} should be taken in a somewhat qualitative sense, although their relative sizes in different cases will presumably correlate the ease of dislocation nucleation in those cases.

$(2\bar{2}1)_{Cu}$ *interface, [1$\bar{1}$4] cracking direction (right-hand crack tip in fig. 9a).* The two slip planes which pass through the crack front are (1$\bar{1}$1) and (1$\bar{1}\bar{1}$), with $\theta = 15.8°$ and $125.3°$, respectively. Due to the large negative phase angle, -79°, for the crack tip field, shear stresses on only the (1$\bar{1}\bar{1}$) system are favorable for nucleation and we find from (3.11) that $G_{disl} = 0.86$ J/m^2.

$(2\bar{2}1)_{Cu}$ *interface, [$\bar{1}$14] cracking direction (left-hand crack tip in fig. 9a).* The (1$\bar{1}\bar{1}$) and (1$\bar{1}$1) planes are at $\theta = 54.7°$ and $164.2°$, respectively, relative to the crack direction. The former is not favorably stressed for nucleation in the -79° near-tip loading phase, and the latter is favorably but only poorly stressed. Thus nucleation occurs on the (1$\bar{1}$1) plane, but only at $G_{disl} = 4.9$ J/m^2.

Thus there is a more than factor-of-five difference between the G_{disl} values for the two directions of crack growth along the same $(2\bar{2}1)_{Cu}$ interface. If G_{cleav} for the interface lies in the broad range between 0.86 and 4.9 J/m^2, then dislocation blunting should occur at the right-hand crack tip in fig. 9a, whereas interfacial decohesion, unimpeded by dislocation nucleation, should occur at the left-hand tip. Even if G_{cleav} lies somewhat below that range, one should still expect strong differences in brittleness for the two cracking directions, since the different G_{disl} reflect different intensities of resolved shear stress along near-tip slip planes, which should also be important for the motion of pre-existing dislocations in the crystal lattice.

Thus we have reached the same conclusion for the bimaterial interfaces as that for metal bicrystal boundaries discussed above: It oversimplifies to say that an interface is intrinsically cleavable. The ductile vs. brittle response of an interface is expected theoretically to be crack direction dependent, and that expectation has been supported experimentally by Wang's (1989) results on symmetric [110] tilt Cu bicrystals, with (2$\bar{2}$1) boundary plane, loaded in tension. That (2$\bar{2}$1) boundary of Wang's bicrystal involves the same crystal plane as that just discussed as the interface for the Cu/sapphire layer combination in fig. 9a. Both cases lead to theoretical prediction of strong crack direction dependence. What is also remarkable, however, is that the brittle [1$\bar{1}$4] direction for the tensile-loaded Cu bicrystal is the ductile direction for the Cu/sapphire composite loaded in bending and, conversely, the ductile [$\bar{1}$14] direction for

14

the bicrystal in tension is the brittle direction for the composite in bending! This is all due to the distinctly different loading phase angles at the crack tips.

$(001)_{Cu}$ interface, $[\bar{1}10]$ cracking direction (right-hand crack tip in fig. 9b), or $[1\bar{1}0]$ cracking direction (left-hand crack tip in fig 9b). These two cases give identical G_{disl} because of the symmetry; there is no crack direction dependence for the two directions considered. At both crack tips $\theta = 54.7°$ and $125.3°$ for $[111]$ slip planes through the tip. The first plane is not favorably stressed for nucleation, given the local -79° loading phase, but the latter is and one again finds $G_{disl} = 0.86$ J/m². Thus this interface is expected to be ductile (or comparatively ductile, if G_{cleav} is somewhat below 0.86 J/m²) for both crack directions considered.

We can better understand this case with reference to fig. 10, in which G_{disl} as calculated from (3.11) is shown as a function of the atomic scale phase angle ψ' at the crack tip: the slip system on which the dislocation nucleates is also indicated. (For the $(2\bar{2}1)$ boundary of fig. 9a, a pair of such plots would be needed, one for each growth direction.) Evidently, the bending specimen of fig. 9b, with $\psi' = -79°$, coincides with a broad minimum in G_{disl}; other specimen designs, or other modes of stressing the crack tip, could lead to higher G_{disl} values and hence would be likely to show more brittle response. E. g., the most brittle case is when $\psi' = -15°$, for which case $G_{disl} = 3.5$ J/m². For Cu/sapphire specimens with characteristic dimensions of order 1 mm, that most brittle case occurs when the angle ψ as introduced in (2.4) is $\cong +12°$. Some results for other test specimens are as follows:

$(001)_{Cu}$ interface, like in fig. 9b, but with the Cu and sapphire interchanged. The configuration considered is like in fig. 7a, with 1 denoting Cu and 2 sapphire. One finds $\psi = 1°$ and thus $\psi' = 42° - 27° = 15°$, which corresponds to $G_{disl} = 1.42$ J/m².

$(001)_{Cu}$ interface, along which a 1 mm sapphire layer is bonded to a 1.5 mm Cu crystal layer; wedging forces act normal to the surface of a long crack between them. The different layer thicknesses are chosen to match bending resistance. This case has a phase $\psi \cong 0$; thus $\psi' \cong -27°$ and $G_{disl} \cong 1.94$ J/m²; it should be much more brittle than the other specimens considered.

$(001)_{Cu}$ interface as in fig. 9b; microcrack ahead of main crack as in fig. 9c. The microcrack is treated for simplicity as a tunnel crack with center at distance r_m ahead of the macrocrack tip; its end B is closest to the macrocrack tip and end A grows in the same direction as the macrocrack. The important factor is that the atomic scale phase angle ψ' is different at the two ends A and B of the microcrack, so G_{disl} is different too. We estimate ψ' at B by regarding the microcrack as a crack of length 2a under a remotely uniform stress field, like in eq.(2.9) and fig. 3; that remote stress field is equated to the local stress which would have acted at distance r_m ahead of the macrocrack tip (i. e., at the center of the microcrack) if the microcrack were not present. Thus, since the concern is with tip B, the

$\sigma_{yy}^{\infty} + i\, \sigma_{yx}^{\infty} \left(= T e^{i\omega} \right)$ for use in (2.9) is chosen as $\sigma_{yy} - i\, \sigma_{yx}$ as would exist along the interface at

15

distance r_m ahead of the macrocrack tip if the microcrack were absent. Thus the phase ω of (2.9) $= -\psi + \epsilon$ $ln(h/r_m)$, where $\psi = -52°$ as before from the bending analysis based on fig. 7. Therefore the atomic scale phase ψ' at B is, from (2.9) and (3.10) with L taken as 2a in the latter, $\psi' = \omega + 2\epsilon - \epsilon \, ln$ (2a/b) $= -\psi - \epsilon \, ln$ (2ar_m/hb). Choosing h and b as in (3.13) and assuming $r_m = 10\mu m$ and 2a $= 5\mu m$ gives $\psi' = 52° - 6° = 46°$ and $G_{disl} = 1.19$ J/m^2 (increasing or decreasing both r_m and 2a by a factor of 5 decreases or increases ψ' by about 6° but has little effect because the range is near the broad relative minimum for G_{disl} at positive ψ' in fig. 10). A similar analysis gives $\psi' = -74°$ and $G_{disl} = 0.86$ J/m^2 at end A of the microcrack, close to the result for macrocrack tip C. It fortuitously turns out for the configuration of figs. 9b,c that all crack tips map close to minima in the G_{disl} vs. ψ' relation, but that for end B is about 40% higher which is consistent with more brittle behavior at B.

While we do not pursue it here, it is possible to find configurations for which there is strongly different behavior predicted at the two ends of such a microcrack, as hinted by the variations in fig. 10. In cases for which a substantially higher G_{disl} is found at end B, we may expect that macroscopic fractures could occur by microcracks forming ahead of the main crack and propagating backwards to join up with it.

(001)$_{Cu}$ *interface in specimen of shape like in fig. 9b, but with Cu in the form of a 25 μm film bonded to sapphire on both sides.* The upper part of the bend specimen in fig. 9b as well as the notched layer at the bottom are sapphire in this case; the Cu is present only as a thin sandwich-like film. This configuration is similar to an experimental arrangement by Evans (1989) and coworkers on an Au film between sapphire layers. We may use the analysis from fig. 7 to find the classical K_I and K_{II} at the macrocrack tip when there is no sandwich; their phase angle is -46°. Using the sandwich solution of figure 6, one then determines that the phase $\psi = -39°$ based on characteristic length L equal to the 25 μm sandwiched layer thickness. Thus the atomic scale phase $\psi' = -59°$ at the main crack tip C, and $G_{disl} = 0.93$ J/m^2 there. If we again assume that a microcrack exists ahead of the main crack, like in fig. 9c, the phase angle at distance $r_m = 10 \mu m$ in the absence of that microcrack would then be -41°. Proceeding as above with the same microcrack size one finds $\psi' = 27°$ at microcrack tip B, so that $G_{disl} = 1.21$ J/m^2 there, and $\psi' = -55°$ so that $G_{disl} = 0.97$ J/m^2 at tip A. Again, tip B is predicted to be more brittle than A, although the difference is smaller in this case.

Based on measurement of the contact angle between Al$_2$O$_3$ particles and a Cu matrix at high temperature, it was reported (Nicholas, 1968) that the work of adhesion $W_{ad} = 0.475$ J/m^2. This value may be taken as an approximation to $2\gamma_{int}$ for the Cu/sapphire interface at room temperature. If it is correct, and if there are no strong lattice trapping effects, then our estimates of G_{disl} above suggest G_{cleav} < G_{disl} for even the most ductile Cu orientations considered, and the interface is intrinsically cleavable. Still, the differences in G_{disl} for different cases signal different amounts of shear stress resolved onto near tip slip systems and, as mentioned, G_{disl} is therefore still expected to correlate with the relative brittleness of response.

4. Effect of Solute Segregation on Interfacial Embrittlement

We review briefly studies of impurity segregation at ceramic and metal/ceramic interfaces and then close with discussion of thermodynamic interrelations between adsorption at interfaces and the ideal work $2\gamma_{int}$ of their brittle separation. For the latter discussion, the thermodynamic framework of Rice (1976) and Hirth and Rice (1980), as used recently by Rice and Wang (1989) and Anderson et al. (1989) in analysis of intergranular embrittlement of Fe-base alloys, is stated in a manner appropriate to interfaces between dissimilar solids, e. g., to metal/ceramic interfaces. Impurity segregation may reduce or enhance $2\gamma_{int}$ depending on the relative propensity of segregation to the interface vs. to the pair of surfaces which result after separation. *Normal* segregators locate more abundantly, at a given equilibrating potential, on the pair of surfaces resulting from fracture than on the unstressed interface, and because of that they cause $2\gamma_{int}$ to decrease. *Anomalous* segregators instead locate more abundantly at the unstresssed interface; they increase $2\gamma_{int}$.

4.1 Review of solute segregation at ceramic and metal/ceramic interfaces. Impurity segregation in metals and their alloys has been studied extensively. It can convert a ductile alloy to one which cracks in a brittle manner along grain interfaces and, in some cases, can ductilize an alloy which is otherwise intergranularly brittle. The scope for impurity-induced alteration is presumably less broad in the more inherently brittle ceramic and metal/ceramic systems, and perhaps for that reason the phenomenon is less studied in those cases.

One well investigated case is Ca segregation at grain boundaries in MgO-doped or NiO-doped Al_2O_3. Stein and Johnson (1975) and Marcus and Fine (1972) observed that the Ca^{+2} ion segregated to the grain boundaries, but Ni and Mg did not. The driving force was argued to be the elastic strain energy produced by the size misfit between Ca^{+2} and Al^{+3} cations. The Langmuir-McLean type isotherm was found to be valid for the Ca^{+2} segregation and a segregation free energy of $\Delta g_b^0 = -121$ kJ/mole in the temperature range 1700 - 1900°C was reported. Jupp and Smith (1980) observed that segregation of CaO at Al_2O_3 grain boundaries reduces the fracture toughness and promotes intergranular fracture, and Funkenbusch and Smith (1975) suggested that Ca^{+2} segregation reduces the cohesive strength of Al_2O_3 grain boundaries. Interfacial embrittlement was also observed in SiC with Al_2O_3 as segregant (Tajima and Kingery 1982) and SiC fiber reinforced lithium-aluminosilicate glasses and glass-ceramics with Nb_2O_3 as segregant (Brennan 1986).

The mechanisms of embrittlement of ceramics by segregation sometimes involves forming an interfacial phase. Faber et al (1988) found that LiF additions to MgO results in a grain boundary phase which causes embrittlement, but this does not seem to apply for Ca^{+2} segregation in Al_2O_3 since no second phase is found. In order to understand the effect of atomic level segregation on the cohesive strength of interfaces, attention must be paid also to the energetics of the segregant on surfaces of the type formed by fracture. The interfacial fracture process in brittle materials is merely a process of separating the interface to produce two free surfaces, and the thermodynamic formalism to follow makes it evident that

17

segregant properties in the final state (the two separated free surfaces) are equally as important as are those of the initial state (the impurity segregated interface) to understanding the phenomenon. There seems to be rather little work done on surface segregation for ceramic materials.

Fischmeister et al (1972) noted that alloying additions could have a marked effect on the contact angle between oxide particles and liquid metals, and thus on the work of adhesion, and that a similar effect of alloying on interfacial energy would be likely to occur in the solid state. They showed that particle/matrix bonding in the Al_2O_3 dispersion strengthened Ni-base and Fe-base alloys was largely dependent on the magnitude of the interfacial energy between the oxide and the metal, which in turn was dependent on matrix alloying. The affects could be positive or negative. It is likely that alloying additions affect the bonding energy at least in part through their effect on impurity segregation behavior.

Segregation at the interfaces between metals and scales, which form on the metal surface during high temperature exposure in an oxidizing atmosphere, received great attention in the development of high temperature alloys. Bonding between Cr_2O_3 and Al_2O_3 scales and metals is basically strong and it is weakened by S segregation at the interfaces. Reactive elements such as Y, Hf and Zr, and dispersions of their oxides improve bonding by preventing S from segregating to the interfaces (Melas et.al. 1988, Smialek, 1987, Smeggil, 1987 and Funkenbusch et al 1985). Recently, Stott (1988) reviewed scale/metal bond strengths and the methods of improving adherence. Besides S, small quantities of Cu, Sn, As, Sb or P can segregate to the scale/metal interface, giving a reduced scale adherence, while Y and Zr can enhance the scale adherence. This enhancement of the adherence may be due to the eliminating of the impurity segregation or due to the segregation of the reactive elements themselves to the oxide/metal interface.

Intergranular fracture of homogeneous materials produces two free surfaces of the same type (although of generally different crystal faces) between which the segregated atoms, or ions in the case of ceramics, can be assumed to divide approximately equally. Debonding of metal/ceramic interfaces produces a metal surface and a ceramic surface, and the thermodynamic properties and distributions of the segregant on these surfaces will in general be different from each other. As mentioned, understanding this final state of interfacial separation is extremely important in understanding the embrittling (or ductilizing) effects of segregation.

4.2 Thermodynamics of interfacial separation. With reference to fig. 11, we focus on the interface as a thermodynamic system which is assumed to be in local equilibrium but which may be (and typically is, at low temperatures) out of composition equilibrium with adjoining bulk solid phases, both before and after separation. Interfacial thermodynamic quantities are defined as Gibbs-like excesses relative to those of the two adjoining phases.

18

We deal exclusively with solutes which are far more abundant along the interface than in a few atomic layers, well removed from the interface, in either adjoining phase. In that case Γ^i, the concentration of segregant i per unit area of interface, is well defined and we assume that an equilibrating chemical potential μ^i can be associated with each segregating species i as it exists at the interface. This is illustrated for a single segregant, represented by black dots, in figs. 11 and 12.

Thus, where u and s are the excesses of energy and entropy per unit area of interface, in a formulation which treats the interface region as being constrained against the necessity of solute composition equilibrium with the adjoining solid phases, one has

$$du = Tds + \sigma_{yy}d\delta_y + \sigma_{yx}d\delta_x + \sigma_{yz}d\delta_z + \sum_i \mu^i d\Gamma^i \qquad (4.1)$$

for reversible changes of state.

The key assumption, embedded in the above equation, is that all thermodynamic functions referring to the interface, e.g., u, s, the $\sigma_{y\alpha}$ and the μ^i, are determined by the values of T, the δ_α and the Γ^i, regardless of the solute concentrations (say, $c_1, c_2 ...$) in the adjoining phases. If (4.1) holds throughout a separation process (one in which δ_y is increased toward indefinitely large values) as assumed in the present modelling, then any distinction is neglected between the pair of free surfaces resulting from such a decohesion fracture and free surfaces having the same T and Γ^i, but produced by a different thermal/mechanical route. The latter route might, e.g., have involved equilibrium solute segregation to the free surfaces at a higher temperature and, possibly, a surface reconstruction.

The work of separation is

$$2\gamma_{int} = \int_{initial}^{final} (\sigma_{yy}d\delta_y + \sigma_{yx}d\delta_x + \sigma_{yz}d\delta_z) \qquad (4.2)$$

where the *initial* state is an unstressed interface (all the $\sigma_{y\alpha} = 0$) and the *final* state is a pair of fully separated surfaces.

Note that in the initial and final states the properties of the interface (or of the pair of free surfaces that it became) depend only on T and the equilibrating potentials μ^i of the segregants. Note that for consistency in deriving consequences of eq. (4.1), one must assume that the segregants distribute between the two separated surfaces, 1 and 2, so that those surfaces are in composition equilibrium with one another (i. e., both surfaces are equilibrated by the same set of μ^i); that is because the formulation assumes that local equilibrium holds within the interface region during separation. Thus the Γ^i in the final state always are to be understood as the total amount i per unit area associated with both fracture surfaces, its distribution between the two being understood to be however necessary as to give identical equilibrating μ^i for both surfaces. The issue did not come up explicitly in previous applications of the formalism of Rice

(1976) and Hirth and Rice (1980) to fracture of interfaces between identical solids, although in that case the equilibrium requirement is met implicitly through assuming that the solutes distribute equally between the pair of surfaces created by the fracture. Of course, the dynamics of actual interfacial fractures, especially those which occur rapidly, may sometimes be such that the two fracture surfaces are well out of composition equilibrium with one another. However, following an argument by Rice (1987) based on the second thermodynamic law, one may show that any such separation process requires more work than what we calculate as the work of separation by assuming composition equilibrium between the two surfaces.

When dealing with initial and final states we use the following system of subscripts: $1/2$ denotes the unstressed interface between solids 1 and 2; $s,1$ denotes the surface of solid 1 after fracture and $s,2$ denotes the surface of 2 after fracture; when we refer to a property of the pair of surfaces after fracture we use subscript fin (for $final$ state). Thus the adsorption behavior of the interface at fixed T is described with the notations $\Gamma^i = \Gamma^i_{1/2}(\mu^1, \mu^2, \mu^3,...)$ for $i = 1,2,3,...$, with inverses $\mu^i = \mu^i_{1/2}(\Gamma^1, \Gamma^2, \Gamma^3,...)$. Such a relation for a single segregant of amount Γ and equilibrating potential μ is sketched in fig. 12 as the curve $\Gamma = \Gamma_{1/2}(\mu)$. The similar relations for the pair of surfaces resulting after fracture are $\Gamma^i = \Gamma^i_{fin}(\mu^1, \mu^2, \mu^3,...)$ and their inverses are $\mu^i = \mu^i_{fin}(\Gamma^1, \Gamma^2, \Gamma^3,...)$. Given that each μ^i is the same for both the fracture surfaces, the functions Γ^i_{fin} can be calculated directly from the adsorption isotherms for the surfaces taken individually. Letting Γ^i_1 denote the amount of i on the fracture surface of solid 1, the adsorption isotherms for that surface have the form $\Gamma^i_1 = \Gamma^i_{s,1}(\mu^1, \mu^2, \mu^3,...)$, $i = 1,2,3,...$, and similarly for surface 2.. Thus noting that $\Gamma^i = \Gamma^i_1 + \Gamma^i_2$, when referring to a pair of surfaces, we see that the functions Γ^i_{fin} are given by

$$\Gamma^i_{fin} = \Gamma^i_{s,1}(\mu^1, \mu^2, \mu^3,...) + \Gamma^i_{s,2}(\mu^1, \mu^2, \mu^3,...) \tag{4.3}$$

The adsorption isotherm in fig. 12 corresponding to the pair of free surfaces has been labelled accordingly, that is, the function $\Gamma = \Gamma_{fin}(\mu)$ marking that isotherm has been written as $\Gamma = \Gamma_{s,1}(\mu) + \Gamma_{s,2}(\mu)$.

The work expression of (4.2) is assured by eq. (4.1) to be independent of path in the δ_α space in the special cases when s (or T) and the Γ^i (or the μ^i) are held constant, although we noted earlier the possibilities that irreversibilities could occur, as local instabilities equivalent to sudden introduction of dislocations along the interface, when shear accompanies opening as inevitably so for elastically dissimilar materials. More generally, on the reversible paths which we assume, the integral does not define a unique value until one characterizes the variation (if any) of T and the Γ^i with the δ_α during separation. We regard T as constant during separation here and consider, like in Rice (1976) and Hirth and Rice (1980), the two cases which follow.

Separation at fixed solute composition. In the first case the Γ^i are all assumed to remain fixed in amount during separation , which is the normal case at low T and with non-mobile segregants

$$(2\gamma_{int})_{[\Gamma] = const} = f_{fin}(\{\Gamma\}) - f_{1/2}(\{\Gamma\}) = f_{s,1}(\{\Gamma_1\}) + f_{s,2}(\{\Gamma_2\}) - f_{1/2}(\{\Gamma\}) \tag{4.4}$$

Here $f = u - Ts$ is the excess Helmholtz free energy per unit area and, in the latter form it is to be remembered that $\{\Gamma\}$ is fixed so that $\{\Gamma_1\}$ and $\{\Gamma_2\}$ must satisfy $\{\Gamma\} = \{\Gamma_1\} + \{\Gamma_2\}$ and assure that $\mu^i_{s,1} = \mu^i_{s,2}$ for each segregant. (The $\{\Gamma\}$ is a shortened notation for the set $\Gamma^1, \Gamma^2, \Gamma^3, \dots$.) Thus, following Rice (1976), or more simply Hirth and Rice (1980) in considering separate-desorb-rejoin-adsorb cycles, it follows from (4.1) that

$$(2\gamma_{int})\{\Gamma\} = const = (2\gamma_{int})_0 - \int_{\{0\}}^{\{\Gamma\}} \sum_i \{ \mu^i_{1/2}(\{\Gamma\}) - \mu^i_{fin}(\{\Gamma\}) \} \, d\Gamma^i \qquad (4.5)$$

where $(2\gamma_{int})_0$ is the work to separate a clean interface (with $\{\Gamma\} = \{0\}$). For the case of a single segregant the reduction of $2\gamma_{int}$ from $(2\gamma_{int})_0$ is marked as an area in fig. 12. Eq. (4.5) links $2\gamma_{int}$ for separation at fixed composition, to quantities which can, in principle, be estimated from solute segregation studies. When the interface and surface coverages are less than values corresponding to full coverage of a set of adsorption sites, idealized as well as having the same low energy relative to solute sites in the bulk, the simple Langmuir-McLean model (McLean, 1957) may be adopted. Thus, considering a single segregant of amount Γ ($= \Gamma_1 + \Gamma_2$ on the pair of fracture surfaces)

$$\mu_{1/2}(\Gamma) = \Delta g^0_{1/2} + RT \ln [\Gamma / (\Gamma^0_{1/2} - \Gamma)]$$
$$\mu_{s,1}(\Gamma_1) = \Delta g^0_{s,1} + RT \ln [\Gamma_1 / (\Gamma^0_{s,1} - \Gamma_1)] \ , \ \ \mu_{s,2}(\Gamma_2) = \Delta g^0_{s,2} + RT \ln [\Gamma_2 / (\Gamma^0_{s,2} - \Gamma_2)] \qquad (4.6)$$

where the (inherently negative) Δg^0 terms are referenced to a bulk phase at the same T, i.e., are based on the expression $\mu = RT \ln [c / (1-c)] = RT \ln c$ for the equilibrating potential when a fraction c of available solute sites are occupied in the bulk, the Γ^0's are the full coverages for the interface and fracture surfaces, respectively. The Δg^0 terms have the form $\Delta h - T\Delta s^0$, where the Δh terms are the enthalpies of segregation (essentially identical to energies of segregation in the present context since pressure times volume terms are negligible for the unstressed boundary and free surface), and the Δs^0 terms are entropies of segregation relating to changes in the atomic vibrational spectrum.

For separations at fixed composition, if the initial coverage (labeled Γ_0 in fig. 12) on the interface falls within the Langmuir-McLean range, then the coverages Γ_1 and Γ_2 on the separated surfaces are also likely to do so and the equations above for the μ can be used directly in the calculation of $2\gamma_{int}$. Following a comparison of the order of different terms like in Rice and Wang (1989), the configurational entropy terms make little contribution to the integral in (4.5) at low T, say, T = 300 K, and thus

$$(2\gamma_{int})\Gamma = \Gamma_0 = (2\gamma_{int})_0 - [\Delta g^0_{1/2} \Gamma_0 - \Delta g^0_{s,1} \Gamma_1 - \Delta g^0_{s,2} \Gamma_2] \qquad (4.7)$$

Here Γ_1 and Γ_2 are determined from the equations $\Gamma_1 + \Gamma_2 = \Gamma_0$ and $\mu_{s,1} = \mu_{s,2}$ in the second line of (4.6). What the latter enforces at low T is that essentially all the solute go to the fracture surface with the smallest (i. e., largest in absolute value) value of Δg^0, at least until the capacity Γ^0 for that surface is reached, and then go to the other surface. Eq.(4.7) is consistent with the results of Smith and Cianciolo (1989), who also consider bimaterial interfaces with impurity segregation.

Data for $\Delta g_{s,1}^0$, where 1 denotes the metal in a metal/ceramic 1/2 interface, can be found in the literature (e.g. for impurities in Fe and Fe alloys, as summarized by Rice and Wang (1989)). Data for ceramic surfaces are limited. It is expected that segregators are normally situated more abundantly on a pair of free surfaces than on an unstressed interface, which means that the combination of terms within the bracket in (4.7) are positive (we remind that the Δg^0's are intrinsically negative), and that the interface and fracture surfaces adsorption isotherms are situated relative to one another like in fig. 12. In this case, $2\gamma_{int} < (2\gamma_{int})0$, the cohesive energy of the interface is reduced by segregation. This might be the case of segregation of S at the metal/oxide interfaces, and is likely also the case for Cu, Sn, As, Sb and P segregation.

Nevertheless, thermodynamics does not preclude the existence of anomalous segregators for which the above inequalities are reversed. Such anomalous segregators increase $2\gamma_{int}$ and hence are expected to increase interfacial toughness. Y and Hf may possibly be examples of anomalous segregators; we mentioned earlier that they enhance the cohesive strength of the metal/oxide interface, although other mechanisms, such as site competition or impurity gettering, are also possible.

Separation at fixed potential of a mobile segregant. While separation at fixed composition seems to be the normal failure mode in low temperature debonding processes, it is useful to consider an opposite limiting case. This limit is separation at constant potential μ ($= \mu_0$, say) implying that there is (for "normal" segregators) solute inflow to the interface region during separation so as to maintain μ fixed, e.g., at the value for an adjoining bulk phases with which there is composition equilibrium. Such conditions require mobility. They are probably met approximately in low temperature hydrogen assisted cracking of some interfaces. They may be met also in high T debonding of metal/scale interfaces.

The $\mu = $ const. $= \mu_0$ path is also shown in fig. 12. The initial state is $\Gamma = \Gamma_0$ and $\mu = \mu_0$. For separations at fixed μ the relevant thermodynamic function is $\gamma = f - \mu\Gamma$, and

$$(2\gamma_{int})\mu = \text{const} = \gamma_{fin}(\mu) - \gamma_{1/2}(\mu) = \gamma_{s,1}(\mu) + \gamma_{s,2}(\mu) - \gamma_{1/2}(\mu) \tag{4.8}$$

Also, following Rice (1976),

$$(2\gamma_{int})\mu = \mu_0 = (2\gamma_{int})0 - \int_{-\infty}^{\mu_0} [\Gamma_{s,1}(\mu) + \Gamma_{s,2}(\mu) - \Gamma_{1/2}(\mu)]\, d\mu \tag{4.9}$$

This result for the alteration of $2\gamma_{int}$ can be interpreted graphically as the sum of the two areas marked in fig. 12. Segregant coverages on a single metal and ceramic free surfaces, and on an interface, at equilibrating potential μ, can be calculated from (4.6) in the range where the Langmuir-McLean model applies.

The difference in $2\gamma_{int}$ at constant Γ vs. that at constant μ can be numerically significant. Examples given by Anderson et al. (1989) for H and (at higher temperature) S on Fe grain boundaries show that

substantial reductions of $2\gamma_{int}$, of order 0.5 $(2\gamma_{int})_0$, can occur in separations at constant μ, under potentials μ that are so low that they equilibrate negligible Γ on the unstressed interface, and hence would involve $2\gamma_{int} \cong (2\gamma_{int})_0$ in rapid separation at constant Γ. This shows that mobility (or slowness of the attempted separation) is a factor which worsens the already deleterious effect of a normal segregator on $2\gamma_{int}$. It also reduces the beneficial effect of an anomalous one. We emphasize again that embrittlement or toughening, at least as it mirrors $2\gamma_{int}$, is always seen to depend on differences between segregant effects on the initial interface and on the two free surfaces created by the fracture. Thus a focus on the electronic alterations induced by segregants in interface, without corresponding study of what they induce on free surfaces, is unlikely to be definitive in explaining solute embrittlement. The differences in energy and entropy between the two states are of primary interest.

Acknowledgement

References

Anderson, P. M. (1986), *Ductile and brittle crack tip response*, Ph. D. Thesis, Harvard University

Anderson, P. M. (1988), *J. Appl. Mech.*, **55**, 814.

Anderson, P. M., and J. R. Rice (1986), *Scripta Metall.*, **20**, 1967.

Anderson, P. M., J.-S. Wang and J. R. Rice (1989), "Thermodynamic and mechanical models of interfacial embrittlement", in *Proc. Sagamore Conf. on Innovations in High Strength Steel Technology*, edited by M. Azrin, G.B. Olson and E. S. Wright, in press.

Argon, A.S., V. Gupta, H. S. Landis, J. A.and Cornie (1989), *J. Mater. Sci.*, **24**, 1406-1412.

Bassani, J. L. and J. Qu (1989), *Mat. Sci. and Eng.* **A107**, 177.

Brennan, J. J. (1986), in *Tailoring Multiphase and Composite Ceramics*, edited by R. E. Tressler, G. L. Messing, C. G. Pantano and R. E. Newnham (Plenum Press, New York,), p. 549.

Cao, H.C. and A. G. Evans (1989), "An experimental study of the fracture resistance of bimaterial interface", *Mech. Mater.*, in press

Charalambides, P.G., J. Lund, A. G. Evans and R. M. McMeeking (1989a), *J. Appl. Mech.*, **56**, 77-82.

Charalambides, P.G, H. C. Cao, J. Lund and A. G. Evans (1989b), "Development of a test method for measuring the mixed mode fracture resistance of bimaterial interfaces", *Acta Met.*, to be published.

Comninou, M. (1977), *J. Appl. Mech.* **44**, 631 and 780.

Comninou, M. and D. Schmueser (1979), *J. Appl. Mech.* **46**, 345.

Dundurs, J. (1968), in *Mathematical Theory of Dislocations*, ASME, New York.

Evans, A. G (1989), private communication.

Faber, K. T., and C. C. Hickenbottom (1988), in *Interfacial Structure, Properties and Design*, edited by M. H. Yoo, W. A. T. Clark and C. L. Briant, (MRS Vol.122), p. 475.

Fischmeister, H. F., E. Navara and K. E. Easterling (1972), *Met. Sci. J.* 6, 211.

Funkenbusch, A. W., J. G. Smeggil and N. S. Bornstein (1985), *Metall. Trans.* 16A, 1164.

Gao, H. and J. R. Rice. (1989), *J. Mech. Phys. Solids*, 37, 155-174.

Hirth, J. P., and J. R. Rice (1980), *Metall. Trans.* 11A, 1502.

Hirth, J. P. and J. Lothe (1982), *Theory of Dislocations*, 2nd Ed., John Wiley & Sons, New York.

Hutchinson, J. W.(1989), "Mixed mode fracture mechanics of interfaces", *Scripta Met.*, in press.

Jupp, R. S., and D. W. Smith (1980), *J. Mater. Sci.* 15, 96.

Kelly, A., W. Tyson and A. Cottrell (1967), *Phil. Mag.*, 15, 567.

Liechti, K. M. and E. C. Hanson (1988), *Int. J. Fracture* 36, 199.

Marcus, H. L., and M. E. Fine (1972), *J. Amer. Cer. Soc.* 55, 568.

Mason, D. D. (1979), *Phil. Mag.* 39, 455.

McLean, D. (1957), *Grain Boundaries in Metals*, Oxford University Press, Oxford.

Melas, I., and D. G. Lees (1988), *Mater. Sci. and Tech.* 4, 455.

Nicholas, M. (1968), *J. Mat. Sci.*, 3, 571.

Oh, T.S., R. M. Cannon and R. O. Ritchie (1987), *J. Am. Ceram. Sco.* 70, C-352.

Rice, J. R. (1976), in *Effect of Hydrogen on Behavior of Materials*, edited by A. W. Thompson and I. M. Bernstein, TMS-AIME, p. 455.

Rice, J.R., (1985), in *Fundamentals of Deformation and Fracture* (Eshelby Memorial Symposium), Cambridge University Press, p. 33.

Rice, J. R. (1987) in *Chemistry and Physics of Fracture*, edited by R. M. Latanision and R. H. Jones, Martinus Nijhof Publishers, Dordrecht, p. 23.

Rice, J.R., (1988), *J. Appl. Mech.* 55, 98.

Rice, J. R. and R. Thomson (1974), *Phil. Mag.*, 29, 73.

Rice, J. R., and J.-S. Wang (1989), *Mater. Sci. Eng.* A107, 23.

Smeggil, J. G. (1987), *Mater. Sci. Eng.* 87, 261.

Smialel, J. L. (1987), *Metall. Trans.* 18A, 164.

Smith, J., and T. V. Cianciolo (1989), *Surface Sci.* 210, L229.

Stein, D. F., and W. C. Johnson (1975), *J. Amer. Cer. Soc.* 58, 485.

Stott, S. H. (1988), *Mater. Sci. Technol.* 4, 431

Suga, T., E. Elssner and S. Schmauder (1988), *J. Composite Materials*, 22, 917.

Suo, Z.(1989a), *Mechanics of interface fracture*, Ph. D. Thesis, Harvard University.

Suo, Z.(1989b), "Singularities interacting with interfaces and cracks", *Int. J. Solids. Structures*, in press

Suo, Z. (1989c), "Singularities, interfaces and cracks in dissimilar anisotropic media", *Harvard University Report Mech-137*, submitted to *Proc. Roy. Soc. Lond. A*.

Suo, Z. and J. W. Hutchinson (1989a), "Interface crack between two elastic layers", *Int. J. Fracture*, in press.

Suo, Z. and J. W. Hutchinson (1989b), *Mater. Sci. Eng.* A107, 135.

Tajima, Y., and W. D. Kingery (1982), *J. Mater. Sci.* 17, 2289.

Thomson, R. (1983), in *Atomistics of Fracture*, edited by R. M. Latanision and J. Pickens, Plenum Press, New York, p. 167.

Thomson, R. (1986), in *Solid State Physics*, edited by H. Ehrenreich and D. Turnbull, Vol. 39, p.1.

Wang, J.-S. (1988), in *Interface Structure, Properties and Design*, edited by M. H. Yoo, C. L. Briant and W. A. T. Clark, MRS Vol. 122, p. 367.

Wang, J.-S. (1989) "Fracture behavior of embrittled f.c.c. metal bicrystals and its misorientation dependence (part 1, experimental)", submitted to *Acta Metall.*

Wang, J. - S. and P. M. Anderson (1989), "Fracture behavior of embrittled f.c.c. metal bicrystals and its misorientation dependence (part 2, theory and analysis)", submitted to *Acta Metall.*

Wang, J. - S., P. M. Anderson and J. R. Rice (1987), in *Mechanical Behavior of Materials -V*, edited by M. G. Yan, S. H. Zhang and Z. M. Zheng, Pergamon Press, p. 191

Wu, K.- C. (1989) "Stress intensity factor and energy release rate for interfacial cracks between dissimilar anisotropic materials", to appear in *J. Appl. Mech.*

Appendix: Angular Functions of the Near-Tip Fields for Joined Isotropic Solids

Listed below are angular functions of eq.(2.1) within for material 1 for modes I and II. For material 2, simply change π to $-\pi$ everywhere. The mode III functions are the same as for a homogeneous solid. A derivation using the Muskhelishvili potentials can be found in Rice (1988).

$$\Sigma^I_{rr} = -\frac{\sinh \varepsilon(\pi - \theta)}{\cosh \pi\varepsilon}\cos \frac{3\theta}{2} + \frac{e^{-\varepsilon(\pi-\theta)}}{\cosh \pi\varepsilon}\cos \frac{\theta}{2}(1 + \sin^2\frac{\theta}{2} + \varepsilon\sin \theta)$$

$$\Sigma^I_{\theta\theta} = \frac{\sinh \varepsilon(\pi - \theta)}{\cosh \pi\varepsilon}\cos \frac{3\theta}{2} + \frac{e^{-\varepsilon(\pi-\theta)}}{\cosh \pi\varepsilon}\cos \frac{\theta}{2}(\cos^2\frac{\theta}{2} - \varepsilon\sin \theta)$$

$$\Sigma^I_{r\theta} = \frac{\sinh \varepsilon(\pi - \theta)}{\cosh \pi\varepsilon}\sin \frac{3\theta}{2} + \frac{e^{-\varepsilon(\pi-\theta)}}{\cosh \pi\varepsilon}\sin \frac{\theta}{2}(\cos^2\frac{\theta}{2} - \varepsilon\sin \theta)$$

$$\Sigma^{II}_{rr} = \frac{\cosh \varepsilon(\pi - \theta)}{\cosh \pi\varepsilon}\sin \frac{3\theta}{2} - \frac{e^{-\varepsilon(\pi-\theta)}}{\cosh \pi\varepsilon}\sin \frac{\theta}{2}(1 + \cos^2\frac{\theta}{2} - \varepsilon\sin \theta)$$

$$\Sigma^{II}_{\theta\theta} = -\frac{\cosh \varepsilon(\pi - \theta)}{\cosh \pi\varepsilon}\sin \frac{3\theta}{2} - \frac{e^{-\varepsilon(\pi-\theta)}}{\cosh \pi\varepsilon}\sin \frac{\theta}{2}(\sin^2\frac{\theta}{2} + \varepsilon\sin \theta)$$

$$\Sigma^{II}_{r\theta} = \frac{\cosh \varepsilon(\pi - \theta)}{\cosh \pi\varepsilon}\cos \frac{3\theta}{2} + \frac{e^{-\varepsilon(\pi-\theta)}}{\cosh \pi\varepsilon}\cos \frac{\theta}{2}(\sin^2\frac{\theta}{2} + \varepsilon\sin \theta)$$

Also, $\Sigma_{zr} = \Sigma_{z\theta} = 0$ for these modes, and $\Sigma_{zz} = \nu (\Sigma_{rr} + \Sigma_{\theta\theta})$.

Fig. 1 Notation for an interface crack.
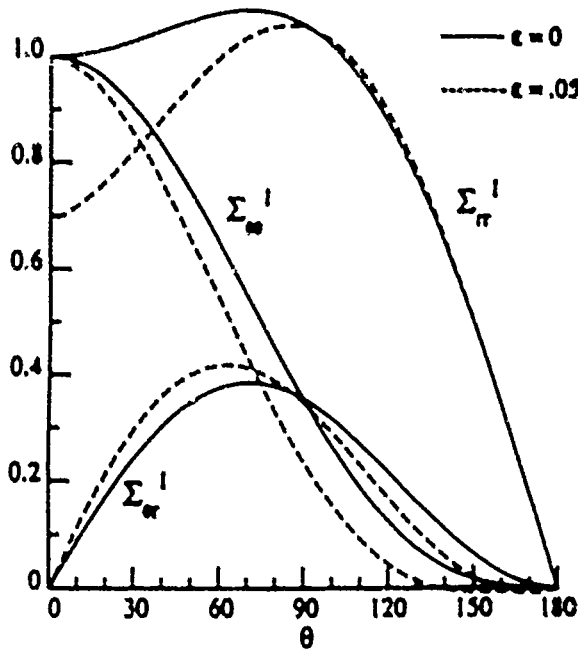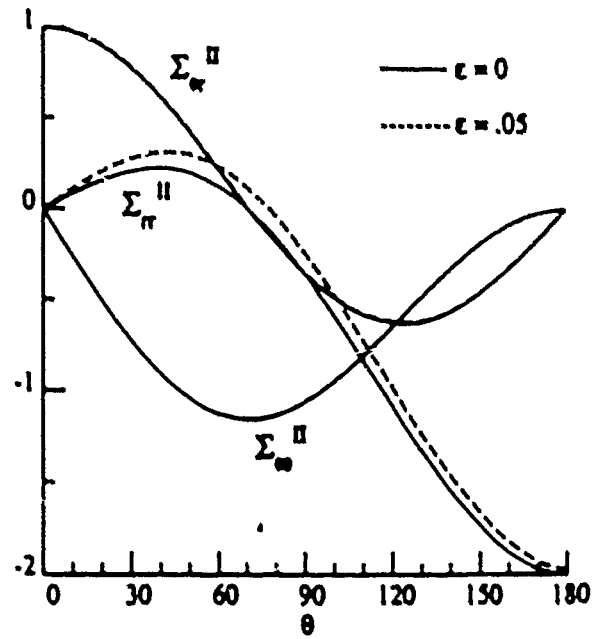

Fig. 2a Angular functions; mode I
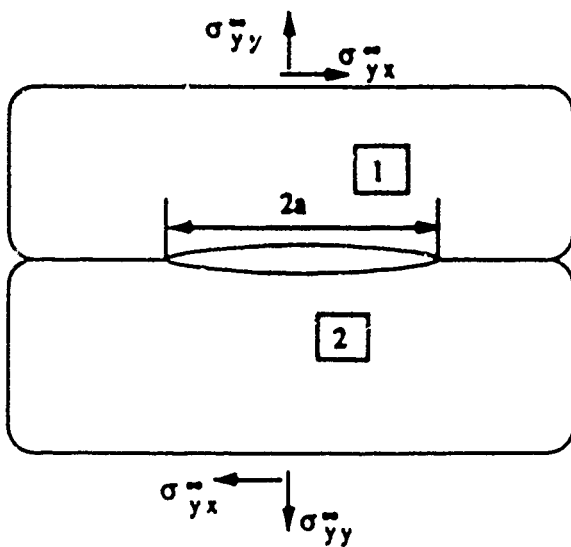

Fig. 2b Angular functions; mode II

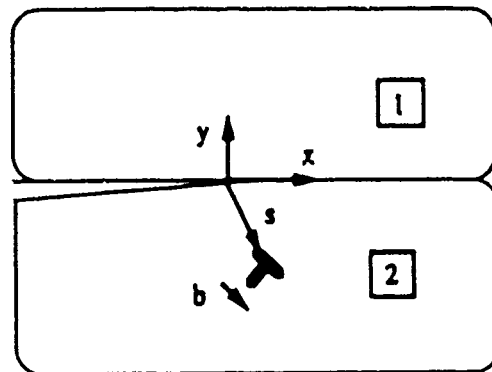Fig. 3 An internal interface crack

Fig. 4 Interaction of a dislocation
with an interface crack tip
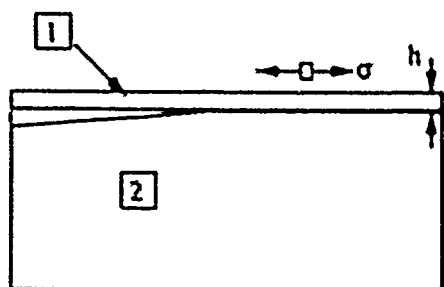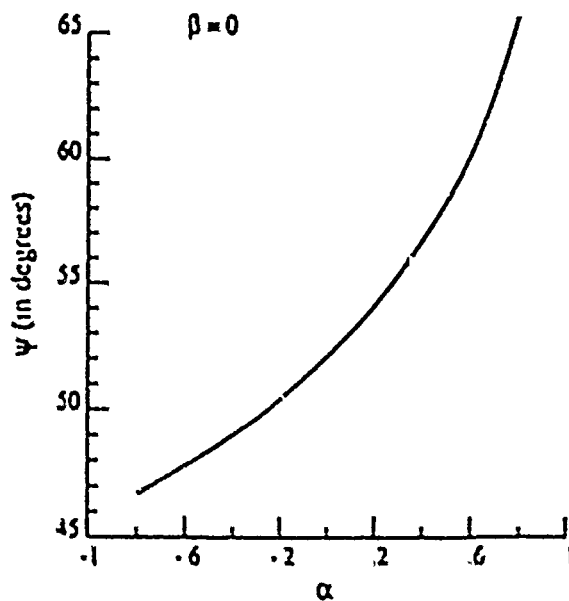
Fig. 5a A thin film on a substrate.

Fig. 5b Phase angle for thin film on substrate
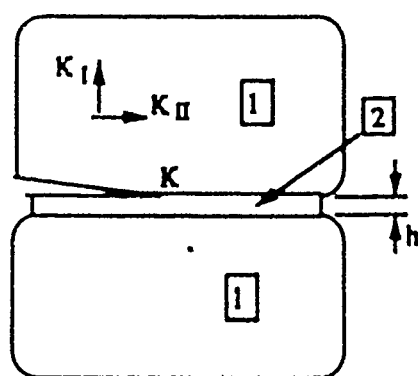

Fig. 6a Sandwich specimens; h is
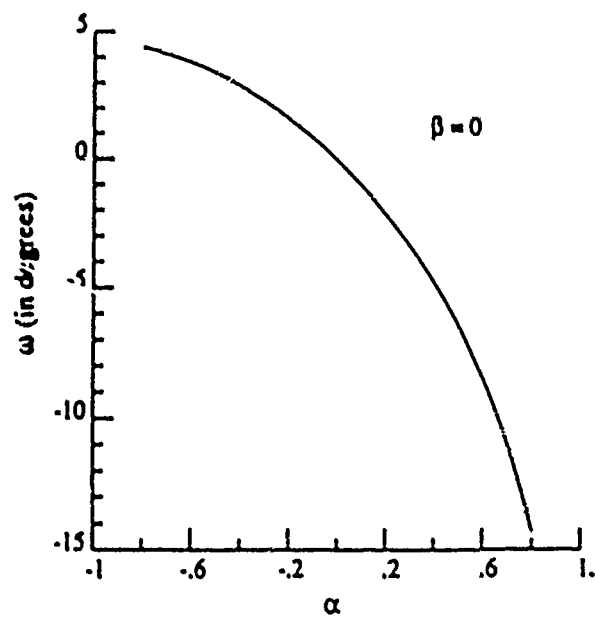small (when h = 0, KI and KII are the
standard stress intensity factors)
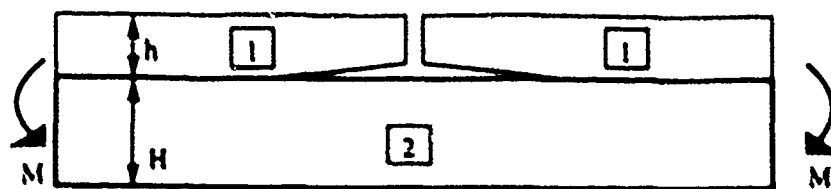

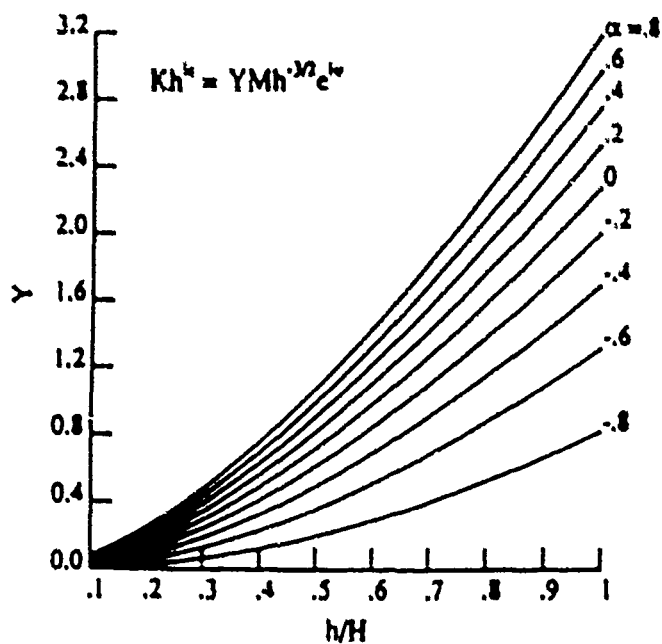Fig. 6b Phase angle shift for sandwiches

Fig. 7a A bending specimen

$Kh^{i\epsilon} = YMh^{-3/2}e^{i\psi}$

Fig. 7b Pre-factor for stress intensity factor

$Kh^{i\epsilon} = YMh^{-3/2}e^{i\psi}$
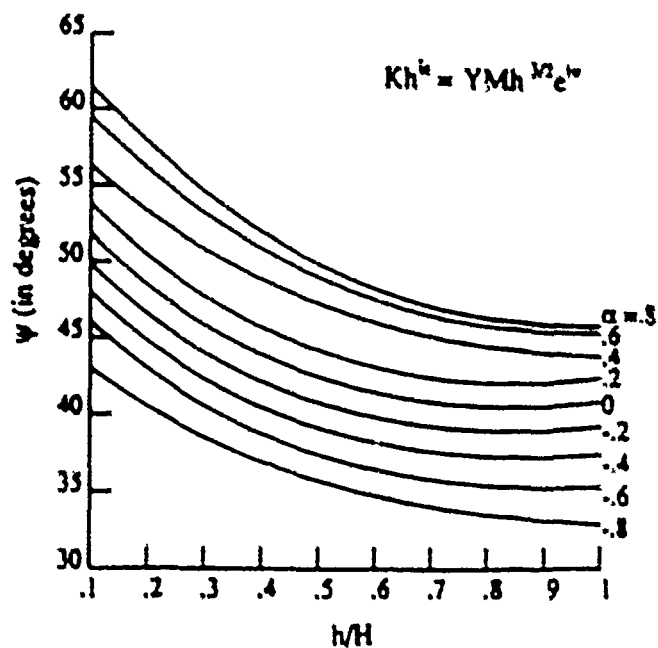
Fig. 7c Phase angle for stress intensity factor
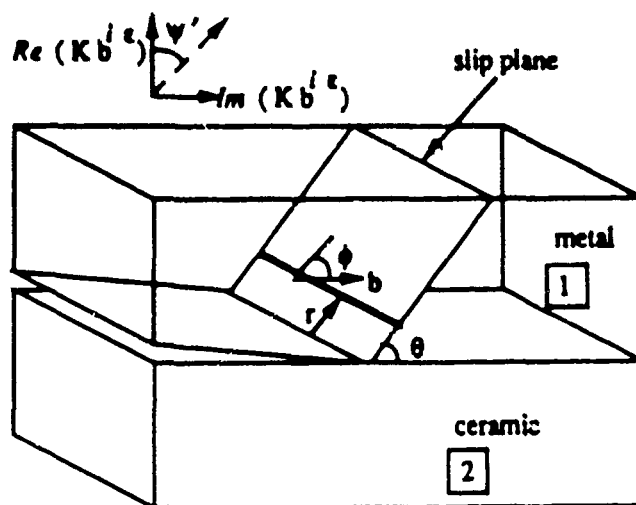
Fig. 8 Dislocation emission from the tip of a metal/ceramic interface crack
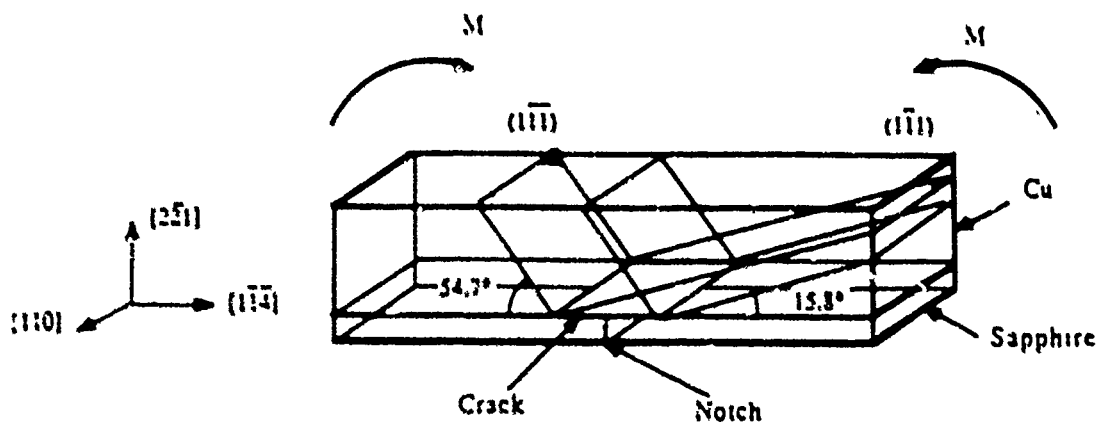
Fig. 9a Cu single crystal with (2$\bar{2}$1) face bonded to sapphire layer; loaded in bending with crack tips along [110]
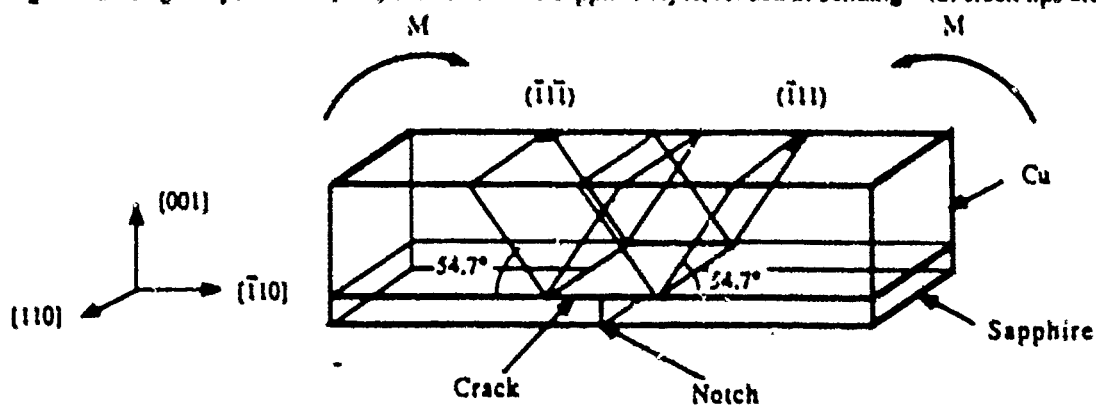
Fig. 9b Cu single crystal with (001) face bonded to sapphire layer; loaded in bending with crack tips along [110]
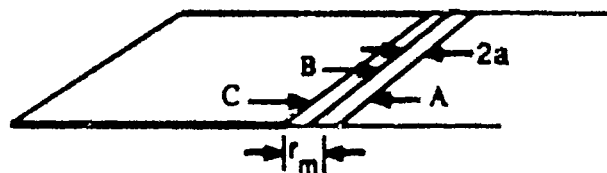
Fig. 9c Microcrack ahead of macrocrack
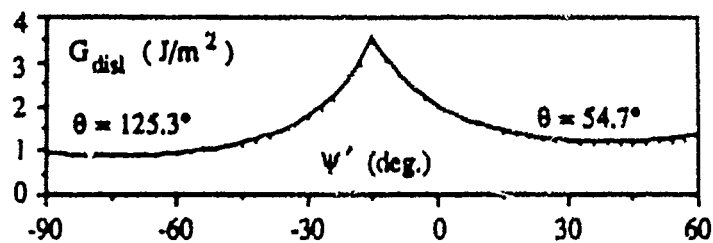
Fig. 10 Dislocation nucleation in Cu; crack tip along [110]; crack plane is (001)
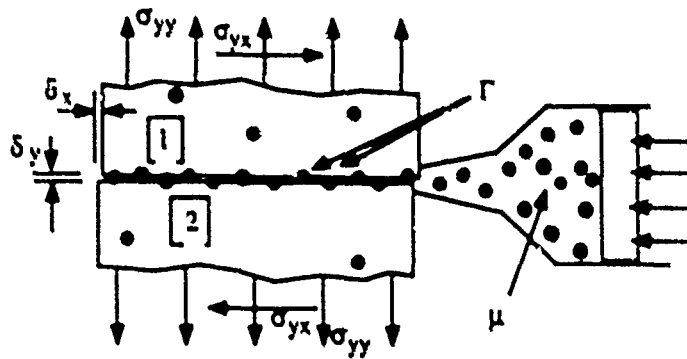
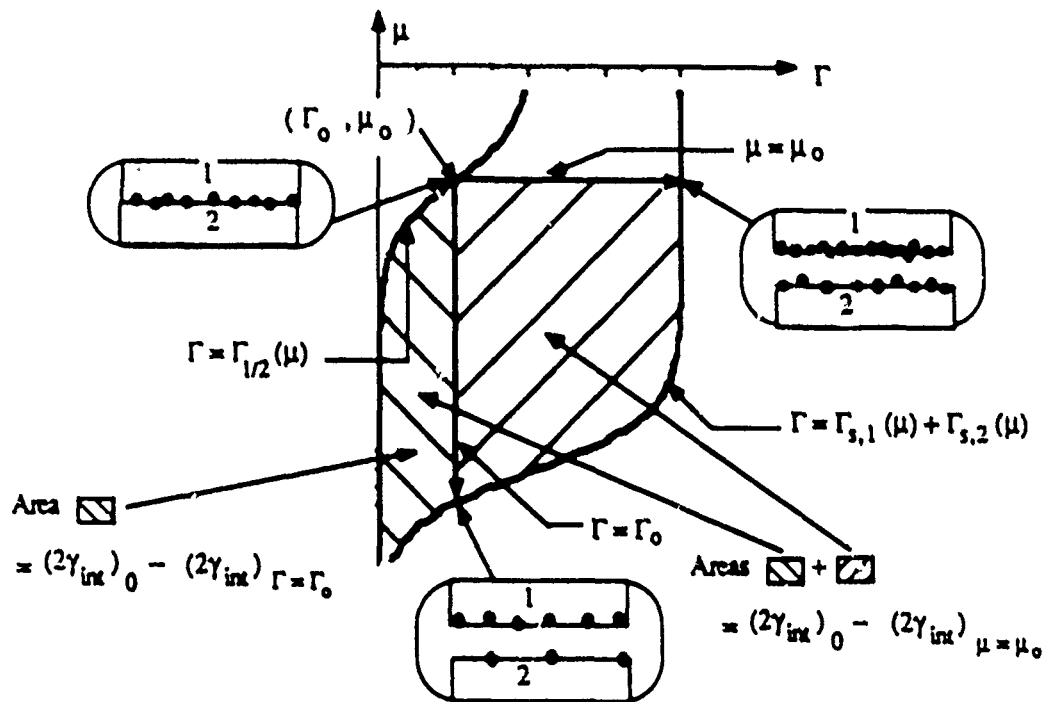Fig. 11 Dissimilar material interface undergoing decohesion; segregated solutes constrained from composition equilibrium with adjoining bulk phases.
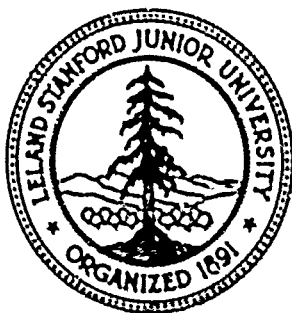

Fig. 12 Use of adsorption isotherms, for unstressed interface and for pair of free surfaces created by separation, to calculate effect of solute adsorption on the ideal work of separation.

Mechanical Properties of Thin Films

William D. Nix

Department of Materials Science and Engineering
Stanford University, Stanford, CA 94305

1988 Institute of Metals Lecture

Mechanical Properties of Thin Films

William D. Nix
Department of Materials Science and Engineering
Stanford University, Stanford, CA 94305

Abstract

The mechanical properties of thin films on substrates are described and studied. It is shown that very large stresses may be present in the thin films that comprise integrated circuits and magnetic disks and that these stresses can cause deformation and fracture to occur. It is argued that the approaches that have proven useful in the study of bulk structural materials can be used to understand the mechanical behavior of thin film materials.

Understanding the mechanical properties of thin films on substrates requires an understanding of the stresses in thin film structures as well as a knowledge of the mechanisms by which thin films deform. The fundamentals of these processes are reviewed. For a crystalline film on a non-deformable substrate, a key problem involves the movement of dislocations in the film. An analysis of this problem provides insight into both the formation of misfit dislocations in epitaxial thin films and the high strengths of thin metal films on substrates. It is demonstrated that the kinetics of dislocation motion at high temperatures are especially important to the understanding of the formation of misfit dislocations in heteroepitaxial structures.

The experimental study of mechanical properties of thin films requires the development and use of non-traditional mechanical testing techniques. Some of the techniques that have been developed recently are described. The measurement of substrate curvature by laser scanning is shown to be an effective way of measuring the biaxial stresses in thin films and studying the biaxial deformation properties at elevated temperatures. Sub-micron indentation testing techniques, which make use of the Nanoindenter, are also reviewed. The mechanical properties that can be studied using this instrument are described, including hardness, elastic modulus and time-dependent deformation properties. Finally, a new testing technique involving the deflection of microbeam samples of thin film materials made by integrated circuit manufacturing methods is described. It is shown that both elastic and plastic properties of thin film materials can be measured using this technique.

Dedication

This paper is dedicated to the memory of Professor G. Marshall Pound, who passed away in May of 1988, during the time this manuscript was being prepared. Professor Pound was Professor of Metallurgy and Materials Science for almost 40 years, having served on the faculties of Carnegie Institute of Technology and Stanford University. He will be remembered for the gusto he brought to classroom teaching, for his vigorous pursuit of excellence in research and for the wisdom he provided in the guidance of students and junior faculty (including the present author). His own Institute of Metals Lecture entitled "Perspectives on Nucleation" was published in this journal in April of 1985. It was a fitting tribute to his professional career and to his scholarly work in the field of Metallurgy and Materials Science. The passing of Professor Pound represents a significant loss for the field of Metallurgy and Materials Science and for the departments at Carnegie-Mellon and Stanford. But our field and our departments are much richer for the 40 year association we had with this great man.

Table of Contents

I.   Introduction

Understanding the relationships between microstructure and mechanical properties has always been one of the primary goals of metallurgy and materials science. Since ancient times, materials have been used primarily for structural, load bearing applications; thus, the mechanical properties of structural materials have always been of paramount importance to our society. This focus on understanding the mechanical properties of structural materials has stimulated a large amount of research that has led to a deep understanding of the microscopic processes responsible for the mechanical behavior of materials. It has also led to the development of an impressive array of advanced, high performance structural materials.

In recent years, more and more attention has been devoted to materials that are not intended for use in load bearing applications. These are the "high tech" materials that have provided the basis for the information revolution. Too frequently, however, we think of these materials soley in terms of their electronic, magnetic or optical properties. We need to remind ourselves that other, non-electronic, properties of these materials can be equally important. For example, the materials used to fabricate microelectronic integrated circuits and magnetic discs must perform their electronic and magnetic functions, but they must also have certain chemical and mechanical properties to be able to do this. These devices must be reliable; they must have structural integrity, and they must retain that integrity over their lifetime; corrosive and mechanical failures must not occur. Thus, these materials, though not selected exclusively for their mechanical and chemical properties, must provide adequate resistance to the mechanical and chemical forces that arise in these applications.

Most of the materials and devices that have led to the information revolution are in the form of thin films deposited on rigid substrates. Thin films ranging in thickness from a few nanometers to a few micrometers, for instance, comprise the most important parts of integrated circuits and magnetic disks. As discussed in the next section of this paper, these films are frequently subjected to very large stresses which can cause a wide variety of deformation and fracture processes to occur. For these reasons, the mechanical properties of thin films, especially those thin films that are used in integrated circuits and magnetic disks, deserve the same kind of study that the mechanical properties of bulk structural materials have received. Just as for bulk structural materials, it is important to understand the microscopic processes responsible for deformation and fracture of these thin film materials, so that the mechanical properties of these materials can be changed through the control of microstructure. These thoughts have motivated much of the work described in this paper.

The viewpoint of this paper is that the field of metallurgy and materials science has the primary responsibility for understanding the mechanical properties of microelectronic and magnetic thin films. In the past, this work has been done largely by scientists and engineers from other disciplines, mainly out of necessity. By contrast, our own field has played a relatively minor role in this work. It is hoped that this paper will encourage others in our field to apply their knowledge and skills in the area of mechanical properties to the problems and opportunities that exist in microelectronic, magnetic and optical thin film materials.

Following this introduction, we give a brief account of some of the mechanical behavior problems that arise in microelectronic and magnetic thin film materials. This review is purely qualitative and tutorial and provides a technological context for the studies of mechanical properties that are described in the remaining sections of the paper. Following this, a brief description of some of the basic concepts relating to stresses and deformation processes in thin films on substrates are given. This background is provided for those readers who are unfamiliar with the field of thin film mechanical properties.

The remaining sections of the paper deal with selected topics relating to the mechanical properties of thin films on substrates. The subject is much too broad and extensive to attempt a comprehensive review here. Selected topics are presented that relate directly to dislocation processes in thin films and provide understanding of both the formation of misfit dislocations in thin film structures and the strength of thin films on substrates. Because experimental techniques for measuring stresses and mechanical properties in thin films are just now emerging, it seems appropriate to focus some of our attention on those techniques.

We begin this study by examining the problem of misfit dislocation formation in epitaxial thin films on single crystal substrates. This problem is important in its own right because there is a great amount of interest in creating device-quality semiconductor alloys by forming epitaxial thin films on dislocation-free silicon substrates. This subject is also of interest here because it provides a vehicle for understanding how dislocation motion occurs in thin films on substrates which, in turn, leads to an understanding of the high biaxial yield strengths of these films. Our study of the strength of thin films on substrates focuses special attention on the substrate curvature technique for studying these properties. The mechanical properties of thin films can also be studied using sub-micron indentation methods. This technique has become popular in recent years because of the availability of commercial depth-sensing instruments with sufficient resolution to study the properties of very thin films on substrates.

2

In the last section of the paper, we report briefly on the development of a new microbeam deflection technique for studying the mechanical properties of thin film materials. This technique makes use of integrated circuit manufacturing methods to create miniature test samples that can be loaded with an indentation instrument. The creation of special test geometries permits the study of deformation in other than the indentation mode (without the hydrostatic pressure that is a natural part of indentation). This approach may also permit a study of fracture and interfacial decohesion of thin films from their substrates.

II. Stress and Mechanical Behavior Problems in Microelectronic and Magnetic Thin Film Materials

As noted in the introduction, the mechanical properties of thin films have become important in recent years because of the extensive use of these materials in integrated circuits and magnetic disks. In this part of the paper we cite some of the mechanical behavior problems that arise in these applications.

Microelectronic Integrated Circuit Materials

In this section of the paper we examine some of the stress and mechanical behavior problems that arise in integrated circuit structures such as the Complimentary-Metal-Oxide-Semiconductor (CMOS) transistor.

CMOS Transistor Materials

Figure 1 shows a schematic picture of the thin film materials and geometries that are used in CMOS transistor devices. The materials are in thin film form, with thicknesses ranging from a few nanometers to about a micrometer. The complex shapes in these structures are also illustrated. It should be noted that a wide variety of materials with very different physical, thermal and mechanical properties are used to create this structure. The materials include semiconductors (which comprise the active part of the device), metals (which serve as conductors to carry current from one part of the structure to another), thermally grown SiO_2, passivation glasses and other dielectric materials. These latter materials ensure electrical isolation of one part of the structure from another and, in some cases, provide mechanical protection for the underlying, electrically

3

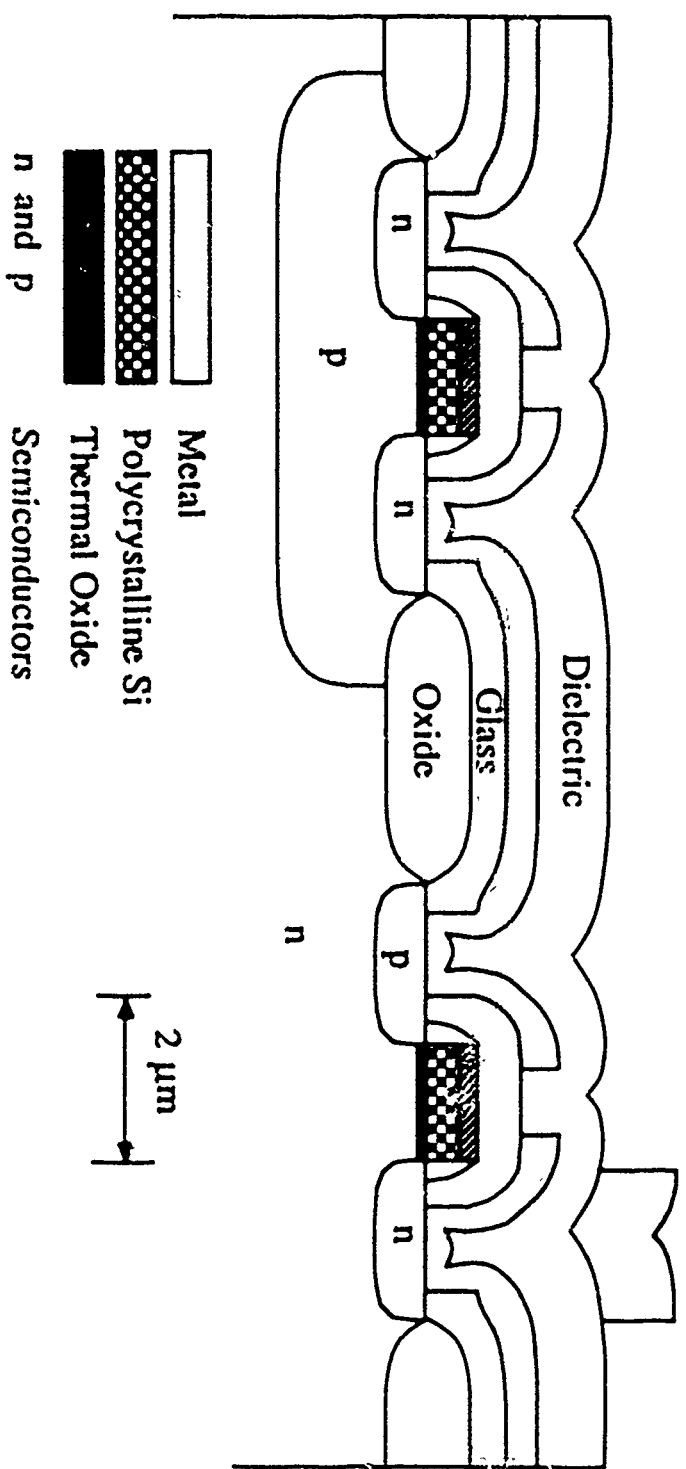# COMPLIMENTARY METAL-OXIDE-SEMICONDUCTOR TRANSISTOR

Figure 1. Cross-sectional structure of a complimentary metal-oxide-semiconductor (CMOS) transistor. The small dimensions of these films and the wide variety of materials involved should be noted. High stresses are present in these films as a result of differences of thermal expansion and of non-equilibrium growth and deposition.

active materials. Note that single crystals, polycrystalline and amorphous phase materials are present.

Stresses in Integrated Circuit Structures

Typically CMOS structures are made by growing or depositing the various thin films onto a substrate of single crystal silicon (a silicon wafer). Because many of these processes occur at elevated temperatures and because some of the materials have differing thermal expansion coefficients, it follows that very large thermal stresses are induced in these materials during manufacture and remain there during the subsequent use of the devices. Thermal stresses of the order of 0.5 GPa are not uncommon. Additional stresses are produced by non-equilibrium growth processes. Most of the growth and deposition processes take place far from thermodynamic equilibrium, with the consequence that highly non-equilibrium microstructures are created. These non-equilibrium microstructures lead to additional stresses caused by the tendency of the film to shrink or expand once it has been deposited onto its substrate. These are sometimes called "intrinsic stresses" in the literature. It is preferable to avoid the implication that bodies can subject themselves to uniform internal stresses and to use the term "growth stresses" instead. Because some of the interfaces are either fully or partially coherent, coherency stresses may also be present in these thin film structures.

Integrated Circuit Failure Processes

For all of the reasons described above, extremely high stresses are commonly present in integrated circuit structures. It follows that these high stresses cause both deformation and, sometimes, fracture, to occur in these materials. It is important to understand the mechanisms that control the mechanical properties of these materials so that integrated circuit structures can be designed for mechanical reliability as well as for electronic device performance.

Even elastic deformation of the thin films in integrated circuit structures can represent a kind of failure. The dimensional tolerances associated with the manufacture of these structures must be very small in order to achieve the small feature sizes needed for high density memory. The patterning of these features must be very precise. Very large thin film stresses can change the dimensions of the silicon wafer during processing, making this high dimensional accuracy difficult to achieve. This would be particularly troublesome if the stresses and distortions were not predictable.

4

Passivation and Interconnect Failures   When plastic deformation occurs in interconnect metals, a variety of unwanted effects can occur. Two of these, shown schematically in Fig. 2, illustrate that inhomogeneous deformation can cause fracture of the passivation to occur, either by dislocation slip or by grain boundary sliding. This fracture, in turn, can cause electrical shorts to occur in the circuit if two metallization layers are allowed to come into contact, or it may lead to some form of delayed failure of the interconnect metal. Such delayed failures might involve either corrosion of the interconnect metal or electromigration (a process of diffusional transport of matter in the interconnect line caused by a very high current density).

One of the purposes of the passivation glass is to constrain the underlying metal layer from deforming by diffusional processes. A crack in the passivation removes the constraint at that location and can allow a net flux of interconnect material to flow to that area, leaving a material deficiency elsewhere in the interconnect line. This has two negative consequences. First the point of material deficiency usually contains voids or cracks that can lead to premature failure of the interconnect. Second, the protrusion of the interconnect material through the crack in the passivation can cause shorts to develop between one metal layer and another. All of these failure processes are initiated by inhomogeneous plastic deformation in the interconnect metal. Needless to say, it is important to understand these deformation processes so that more failure resistant microstructures can be designed.

Substrate Failures   The thin film stresses represent forces that must be balanced by stresses in the substrate. Because the substrates are usually extremely thick compared to the films, the stresses in the substrate are usually quite small and often negligible. This is not always the case, however. In some cases, the compressive stresses in dielectric films can be so great that the corresponding tensile stresses in the uppermost layer of the substrate can cause dislocations to be nucleated there. Even cracking of the substrate can occur, especially if notches are present in the structure. Figure 3 illustrates the conditions that can lead to substrate cracking. Here, the passivation material is in a state of biaxial compression, which causes bending of the substrate to occur. These bending stresses in the substrate are tensile just beneath the passivation film. The stresses at the notch in the substrate can be sufficiently high to either nucleate dislocations there or perhaps to initiate fracture. The compressive passivation in the notch itself also serves to create a high tensile stress in the substrate and this further increases the likelihood of crack initiation. The high aspect ratio trench structures that are now being used in some devices are examples of structures that might be susceptible to this kind of mechanical failure.

FAILURE ASSOCIATED WITH INHOMOGENEOUS

PLASTIC FLOW

PASSIVATION CRACKING
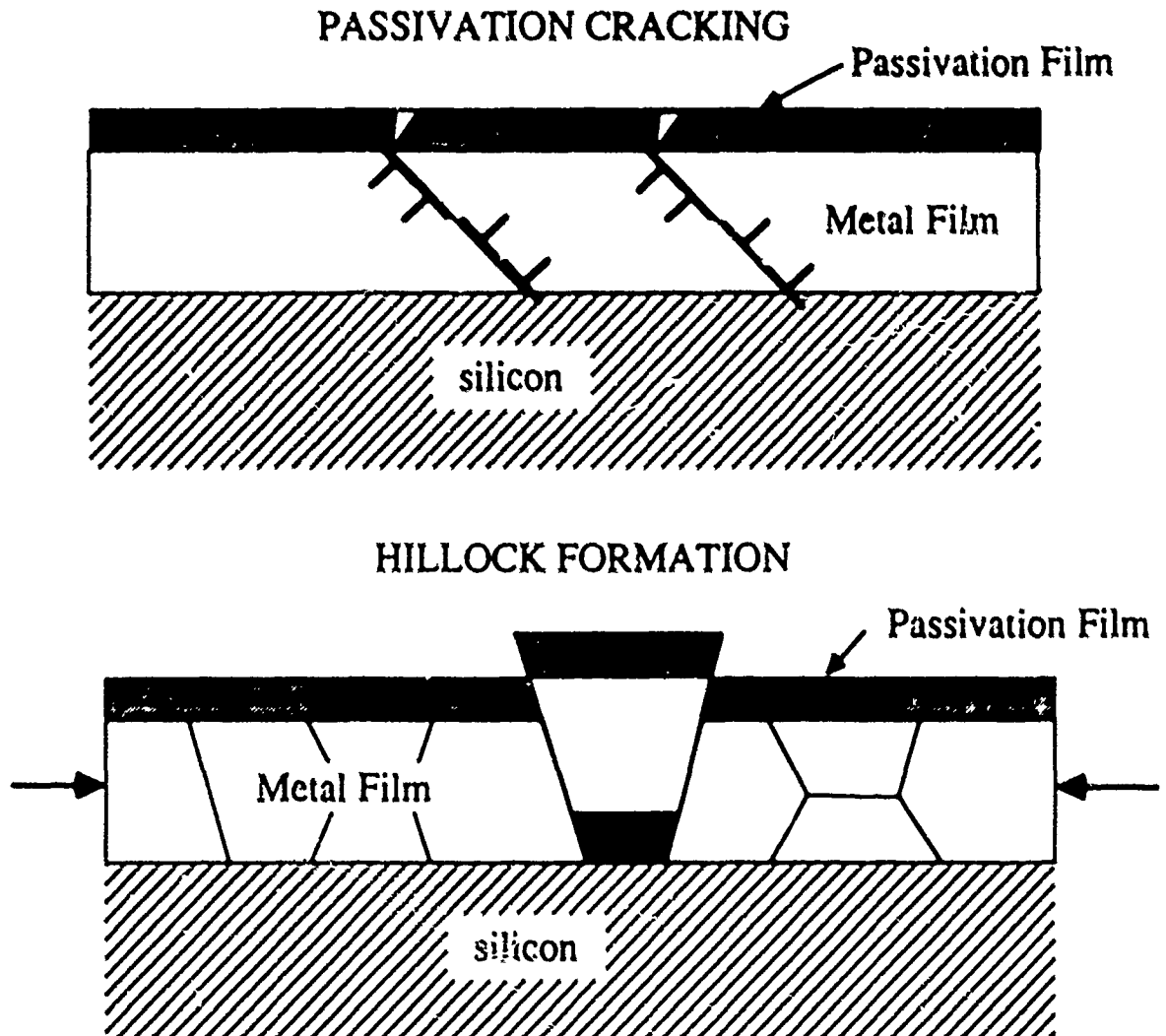
HILLOCK FORMATION

Figure 2.   Mechanisms of inhomogeneous deformation in thin films leading to cracking of the associated passivation film. Dislocation pile-ups can cause passivation cracking and grain boundary sliding can cause hillock formation. Cracking of the passivation film can lead to other failure mechanisms.

SUBSTRATE CRACKING

Compressive Passivation

Substrate

Figure 3. Substrate cracking caused by stresses in a passivation film; the stresses in the substrate are concentrated at notches.

Magnetic Disk Materials

Magnetic hard disks are also composed of thin film structures and thus the mechanical properties of the films involved are of technological importance. Here we describe the materials that are used in these applications and the mechanical forces that can lead to failures.
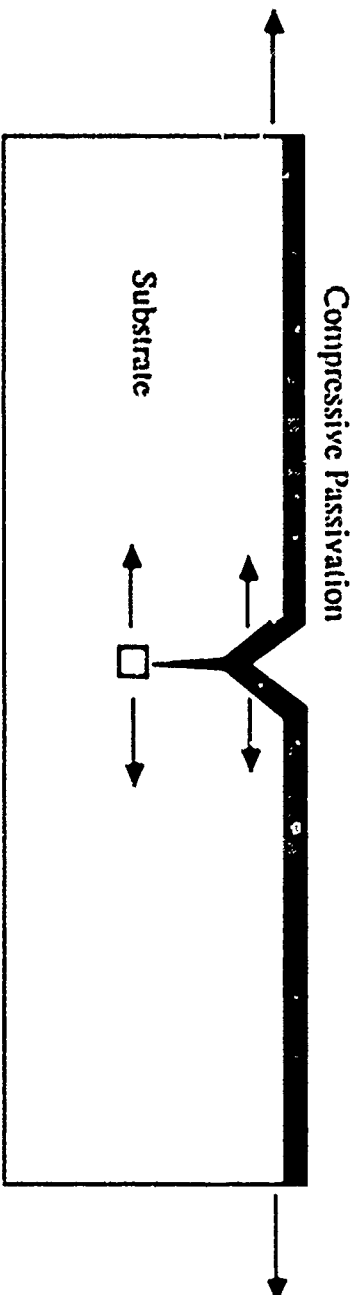
Two different approaches to hard disk construction are shown in Fig. 4. The differences relate primarily to the form of the magnetic media used in the device. In the older, more mature technology, the magnetic media consists of fine magnetic particles dispersed in a resin matrix whereas in the newer technologies, the magnetic material is in the form of a thin magnetic film. The basic operation of the disk is the same, however. In both cases, the disk is rotated at a high speed (with surface velocities as high as 50 mph!) while the head "flies" aerodynamically just above the surface of the disk (within 2000Å in the case of thin film media) and "writes" or "reads" information on the disk. The head "writes" information by magnetizing or demagnetizing a small portion of the magnetic material on the disk. It also "reads" the information by sensing the magnetic field of the magnetized portion.

The dimensions of the materials involved, the differences in their physical properties and the non-equilibrium processes used to fabricate hard disk structures can all lead to stress and mechanical behavior problems that are similar to those that arise in integrated circuit structures. However, the presence of the head, which flies at such a high relative velocity so close to the disk, introduces completely new problems in these devices -- the problems of friction and wear. Frequent collisions between the head and the disk cannot be avoided. The magnetic films must be able to withstand this kind of abrasion without losing any of the information stored there. The head also makes a crash landing on the disk each time the device is turned off and a sliding take-off when the disk is turned on once again. These events also produce friction and wear that may eventually lead to failure.

The purpose of some of the microstructural features in hard disk materials is to provide resistance to friction and wear. The carbon coating on the top of the film is a hard surface with a low coefficient of friction. This permits the head to slide on the surface of the disk without producing mechanical damage. The particles of Al_2O_3 in the fine particle media also provide mechanical durability. During take-off, landing and collision events, the head slides on the tops of the particles, rather than gouging into the softer matrix containing the magnetic particles. The mechanical properties of all of these sub-micron features are crucial to the successful performance

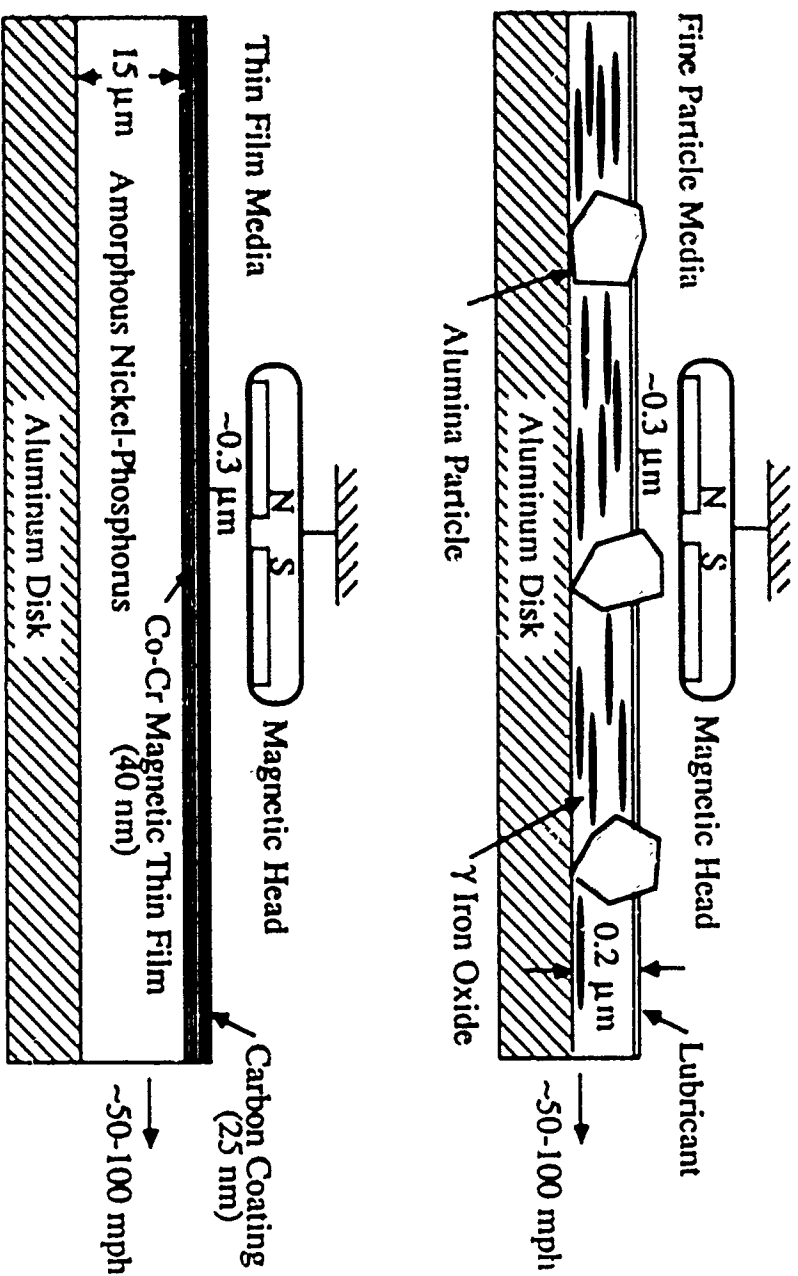# MAGNETIC HEAD-DISK INTERACTIONS

Figure 4. Illustration of magnetic head-disk interactions. Two types of magnetic media are shown: fine-particle media and thin-film media. The head "flies" aerodynamically very close to the disk at a very high velocity. Collisions between the head and disk occur frequently.

of these devices. Needless to say, the measurement and understanding of the mechanical properties of these materials is of great importance.

III. Basic Concepts and Equations Relating to Stresses in Thin Films on Substrates

Because much of this paper relates to processes that are driven by the stresses that are present in thin films on substrates, it is important to have a clear understanding of the stress analysis of the thin film/substrate problem. In this section of the paper we give an elementary treatment of this problem and show how to calculate the stresses in various circumstances and how these stresses lead to elastic bending of the substrate. In a later section of the paper we make use of these relations to show how the stress in the film can be determined from a measurement of the substrate curvature. We present this analysis for those readers who have not studied the basic aspects of this problem. More detailed and in-depth treatments can be found in the literature [1-3].

The Mechanics of Stresses in Thin Films on Substrates

We start by considering a film/substrate composite that is completely free of stress, as shown in Fig. 5. We assume that the film is very thin compared to the substrate (thin film approximation) and we also assume that the lateral dimensions of the film and substrate are much greater than their total thickness. Because the film is under no stress, we can imagine removing it from the substrate and allowing it to stand in a stress free state. In this state, the lateral dimensions of the film will exactly match those of the substrate from which it was removed. If the state of the film is never changed, it can always be reattached to the substrate without causing any stresses to be generated in either the film or substrate. However, if the dimensions of the film change in any way, then elastic strains and stresses will develop when the film is reattached to the substrate. For the present analysis, we consider that the film experiences a uniform volume change (dilatational transformation strain, e_T) in the detached state. This dilatational strain is measured relative to that of the substrate. This is shown in the figure as a negative dilatation (a volume shrinkage). For a pure dilatational strain, the principal strain components are $\varepsilon_{xx}(T) = \varepsilon_{yy}(T) = \varepsilon_{zz}(T) = e_T/3$. There are many ways in which the volume of a film might change relative to the substrate. For example, if the film and substrate have different thermal expansion coefficients a change in temperature will produce the relative volume change discussed here. In addition, the annihilation of excess vacancies, dislocations and grain boundaries are processes that lead to volume changes due to

7

FILM STRESS - SUBSTRATE CURVATURE RELATIONS

Film ➘   Stress Free Film on Substrate

Substrate

Remove Film from Substrate

Allow Film Dimensions to Change Relative to Substrate

Apply Stress to Return Film to Substrate Dimensions

←   →

Reattach Film to Substrate and Remove External Forces
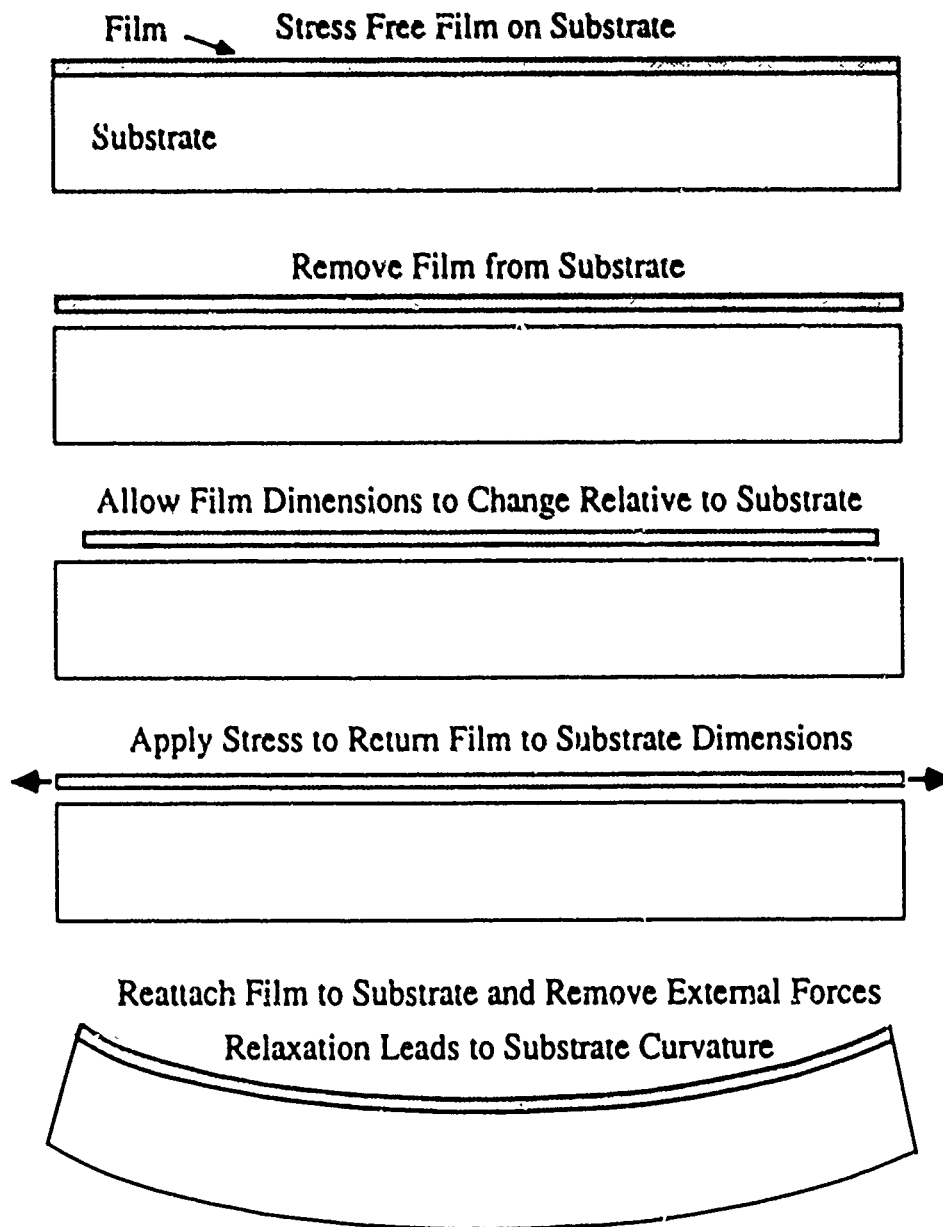Relaxation Leads to Substrate Curvature

Figure 5.   Illustration of the relationship between the biaxial stress in a thin film and the associated bending of the substrate.

densification. Also, phase transformations and composition changes can produce dilatational strains in the film. For the case of heteroepitaxial films, a natural misfit exists between film and substrate in their stress free state.

We now consider the process of reattaching the film to the substrate. Because the lateral dimensions of the film no longer match those of the substrate, a biaxial stress must be imposed on the film to elastically deform it, so that it again fits the dimensions of the substrate. The stress required to do this produces elastic strains ϵ_{xx} and ϵ_{yy} that exactly compensate these components of the transformation strain. Thus

$$c = \epsilon_{xx} = \epsilon_{yy} = -\frac{c_T}{3} \quad . \tag{1}$$

Using Hooke's law, this leads to a biaxial stress in the film

$$\sigma = \sigma_{xx} = \sigma_{yy} = Mc \quad , \tag{2}$$

where M is the biaxial elastic modulus of the film. For isotropic elasticity, the biaxial modulus of the film is simply

$$M = \frac{E}{1-v} \quad , \tag{3}$$

where E is Young's modulus and v is Poisson's ratio.

In Fig. 5 we note that the tensile forces needed to deform the film to match the dimensions of the substrate produce a biaxial stress state in the film. As long as these forces are present, the stress in the film does not change when it is reattached to the substrate. We now consider the changes that occur when these edge forces are removed after the film is again perfectly bonded to the substrate. A principle of superposition can be used to understand what happens when the edge forces are removed. We remove the forces at the edges by superimposing forces of opposite sign on the edges of the film. These additional forces remove the normal tractions from the edges of the film, and they produce shear stresses on the film/substrate interface near the edges of the film. These shear stresses on the interface near the edges of the film provide the forces needed to maintain the biaxial stress in the film. These forces also cause the substrate to bend elastically, as discussed below.

8

Single Crystal Thin Films

For the case of single crystal films or for polycrystalline films with strong crystallographic texture, the anisotropy of the elastic properties must be taken into account. Here we give some results for cubic single crystal films. For a film in which the (001) cube plane lies parallel to the plane of the film, the biaxial elastic modulus is isotropic in the plane of the film and is given by

$$M(001) = C_{11} + C_{12} - \frac{2C_{12}^2}{C_{11}} \quad , \tag{4}$$

where C_{11} and C_{12} are components of the stiffness matrix. We note that neither Young's modulus nor Poisson's ratio alone are isotropic in the (001) plane, but the ratio $E/(1-v)$, which forms the biaxial modulus, is isotropic in that plane. A single crystal film with a (111) plane lying in the plane of the film is fully isotropic in that plane. In this case, the biaxial modulus is

$$M(111) = \frac{6C_{44}(C_{11} + 2C_{12})}{C_{11} + 2C_{12} + 4C_{44}} \quad . \tag{5}$$

For a single crystal film with a (011) plane parallel to the plane of the film, the elastic properties are not isotropic in the plane of the film. The elastic moduli in the mutually perpendicular [100] and [01$\bar{1}$] directions are different. If such a film is subjected to in-plane stresses such that the result is an equal biaxial strain, $\varepsilon = \varepsilon_{xx} = \varepsilon_{yy}$, the stresses in the two directions can be computed using the following moduli:

$$M[100] = C_{11} + C_{12} - \frac{C_{12}(C_{11} + 3C_{12} - 2C_{44})}{(C_{11} + C_{12} + 2C_{44})} \tag{6}$$

$$M[01\bar{1}] = \frac{(2C_{11} + 6C_{12} + 4C_{44})}{4} - \frac{(2C_{11} + 2C_{12} - 4C_{44})}{4} \frac{(C_{11} + 3C_{12} - 2C_{44})}{(C_{11} + C_{12} + 2C_{44})} \tag{7}$$

Substrate Bending and Substrate Stresses

The edge forces exerted on the substrate by the biaxial stress in the film cause the substrate to deform elastically in biaxial bending. The amount of bending depends on the thickness and biaxial elastic modulus of the substrate. If the biaxial elastic modulus is isotropic in the plane of the substrate, as it is for (001) and (111) substrates of cubic materials (eg. silicon), then the film/substrate composite adopts the shape of a spherical shell when it deforms. A simple biaxial bending analysis shows that the curvature, K, and radius of curvature, $R = K^{-1}$, are given by

$$K = \frac{1}{R} = \frac{1}{M_s} \frac{6\sigma_f h_f}{h_s^2} \quad , \tag{8}$$

where $M_s = (E/1-v)_s$ is the biaxial elastic modulus of the substrate, σ_f is the biaxial tensile stress in the film, h_f is the thickness of the film and h_s is the thickness of the substrate. The edge force per unit length is represented by $\sigma_f h_f$ in this expression. We note that this relation can be inverted to give the stress in the film as a function of the substrate curvature caused by that stress, as shown by

$$\sigma_f = M_s \frac{h_s^2}{6 h_f R} \quad . \tag{9}$$

For single crystal substrates, the biaxial elastic modulus, M_s, must be computed from the components of the stiffness matrix for that crystal. The modulus M_s is given by eqns. (4) and (5) for (001) and (111) substrates, respectively. When the (011) plane is parallel to the plane of the substrate (or for non-cubic substrates), the substrate is not elastically isotropic in that plane and does not bend symmetrically.

It should be noted that eqn. (8) for the elastic bending of the substrate does not depend on the elastic properties, or any other mechanical properties of the film. The reason for this relates to the thin film approximation that is made in the derivation of this result. Typically, films about 1 μm thick are deposited onto substrates that may be 500 to 1000 times thicker. In such cases, the flexural modulus of the thin film/substrate composite is completely dominated by the properties of the substrate. The properties of the film have a negligible effect on the bending. It follows that when multiple thin films are deposited sequentially onto a much thicker substrate, each film causes a fixed amount of bending to occur, irrespective of the order in which the films are deposited. The

amount of curvature change is determined by the stress and thickness of each film. The total change of substrate curvature is simply the sum of the curvature changes associated with the presence of each film. Of course, the sign of the curvature change caused by each film must be taken into account. For cases in which the films are not thin compared to the substrate, a more elaborate bending analysis must be used that takes account of the elastic properties of the films [3].

It is commonly believed that the stresses in a thin film of uniform thickness can cause significant stresses to develop in the underlying films and in the substrate. Specifically, it is widely assumed that depositing a thin film of a passivation material like Si_3N_4, which typically is in a state of high biaxial compressive stress, will cause correspondingly large biaxial tensile stresses to develop in the substrate and in underlying thin films. A simple bending analysis using the thin film approximation shows that this is not true. Although the compressive stress in the passivation film is compensated by tensile stresses in the underlying layers, these tensile stresses are usually extremely small and can often be ignored. Typically, the maximum tensile stress induced in the underlying layers, $\sigma_{induced}$, by the compressive stress in the passivation film, $\sigma_{pass.}$, is given by

$$\sigma_{induced} = -\frac{3h_p}{h_s}\sigma_{pass.} \quad . \tag{10}$$

This formula is obtained by balancing the force per unit length in the passivation film, $\sigma_{pass.}h_p$, with an opposing force per unit length and moment per unit length in the substrate. For the typical dimensions of $h_p \approx 1$ μm and $h_s = 500$ μm, this shows that the magnitude of the stress in the underlying layers is less than one percent of the compressive stress in the passivation. For patterned structures, the interaction of stresses in different films can be significant and must be taken into account.

Sources of Strain

As discussed above, the stresses in thin films on substrates can be viewed as arising from the misfit that must be accommodated elastically when the film is attached to the substrate.

Thermal Strains

There are various types of strains that can develop. For thermal mismatch problems, the elastic strain needed to fit the film to the substrate is

11

$$\varepsilon = - (\alpha_f - \alpha_s)(T - T_0) = - \Delta\alpha\Delta T \quad , \tag{11}$$

where α_f and α_s are the linear thermal expansion coefficients of the film and substrate, respectively, T is the current temperature and T_0 is the initial temperature at which the film and substrate were in a stress free state. A typical case is one in which a metal film is deposited onto silicon in a stress free state at high temperatures and subsequently cooled to room temperature. In such cases $\alpha_f > \alpha_s$ and $T < T_0$ so that the elastic accommodation strain is positive and tensile stresses are developed in the film.

Growth Strains

As discussed earlier, if the density of the film changes after it has been bonded to the substrate, an "intrinsic" or growth stress develops in the film. In this case, the elastic accommodation strain is given by

$$\varepsilon = - \frac{e_T}{3} \quad , \tag{12}$$

where e_T is the dilatational "transformation" strain associated with the change in density. A film that densifies when it is attached to the substrate must be subjected to biaxial tensile strain (and corresponding tensile stress) to match the dimensions of the substrate.

Epitaxial Strains

For epitaxial films on thick substrates, the elastic accommodation strain is simply

$$\varepsilon = \frac{\Delta a}{a} = \frac{a_s - a_f}{a_f} \approx \frac{a_s - a_f}{a_s} \quad , \tag{13}$$

where a_f and a_s are the lattice parameters of the film and substrate, respectively. Here all of the elastic accommodation is assumed to take place in the film because the substrate, being so much thicker than the film, is essentially rigid.

IV. Formation of Misfit Dislocations in Epitaxial Thin Films

We begin our study of mechanical properties of thin films on substrates by considering the formation of misfit dislocations in epitaxial thin films on single crystal substrates. This problem is important in its own right because of the technological interest in forming dislocation-free epitaxial thin films or single crystal substrates. It is also important because it provides a basis for understanding the dislocation processes responsible for plastic deformation of thin films on non-deformable substrates.

There is a great amount of interest in growing semiconductor thin films on dislocation free substrates using heteroepitaxy. The basic idea of this technology is that semiconductors that are difficult to grow as bulk crystals might be grown by forming heteroepitaxial layers on dislocation-free single crystal substrates. Of course, any lattice mismatch between the film and substrate must be accommodated by a uniform strain in the film (together with slight bending of the substrate) and this typically leads to very large biaxial stresses in the film. Intuitively, one might expect these large stresses to be relaxed by plastic flow in the film (through dislocation nucleation and motion) regardless of the film thickness. But this cannot occur in very thin films because the energies of the dislocations created by such relaxation processes are greater than the recovery of strain energy associated with the relaxation. Thus, there is a critical film thickness, h_c, below which the film is stable with respect to dislocation formation. Below that critical thickness misfit dislocations cannot and do not form in heteroepitaxial structures.

Equilibrium Theory

The equilibrium theory of misfit dislocation formation was developed by van der Merwe [4] and Matthews [5-7] and is now well established. Figure 6 illustrates the essential features of the theory. As shown in the figure, the energy (per unit film area) of a very thin film is lowest when no dislocations are present. Such films are perfectly coherent with the substrate in their equilibrium state and have an energy (per unit area) given by

$$E_{hom\ ogeneous} = Mh\varepsilon^2 \qquad (h<h_c) , \qquad (14)$$

where M is the biaxial elastic modulus of the film, h is the film thickness and ε is the biaxial elastic strain that must be imposed on the film to bring the lattice of the film into coincidence with that of the substrate. We note that $M\varepsilon$ is the biaxial stress in the film. The fact that the equilibrium state

13

FORMATION OF MISFIT DISLOCATIONS IN EPITAXIAL FILMS

$$E = Mh\varepsilon^2$$

$$E = Mh\left(\varepsilon - \frac{b}{S}\right)^2 + \frac{\mu b^2}{2\pi(1-v)}\frac{2}{S}\ln\left(\frac{\beta h}{b}\right)$$
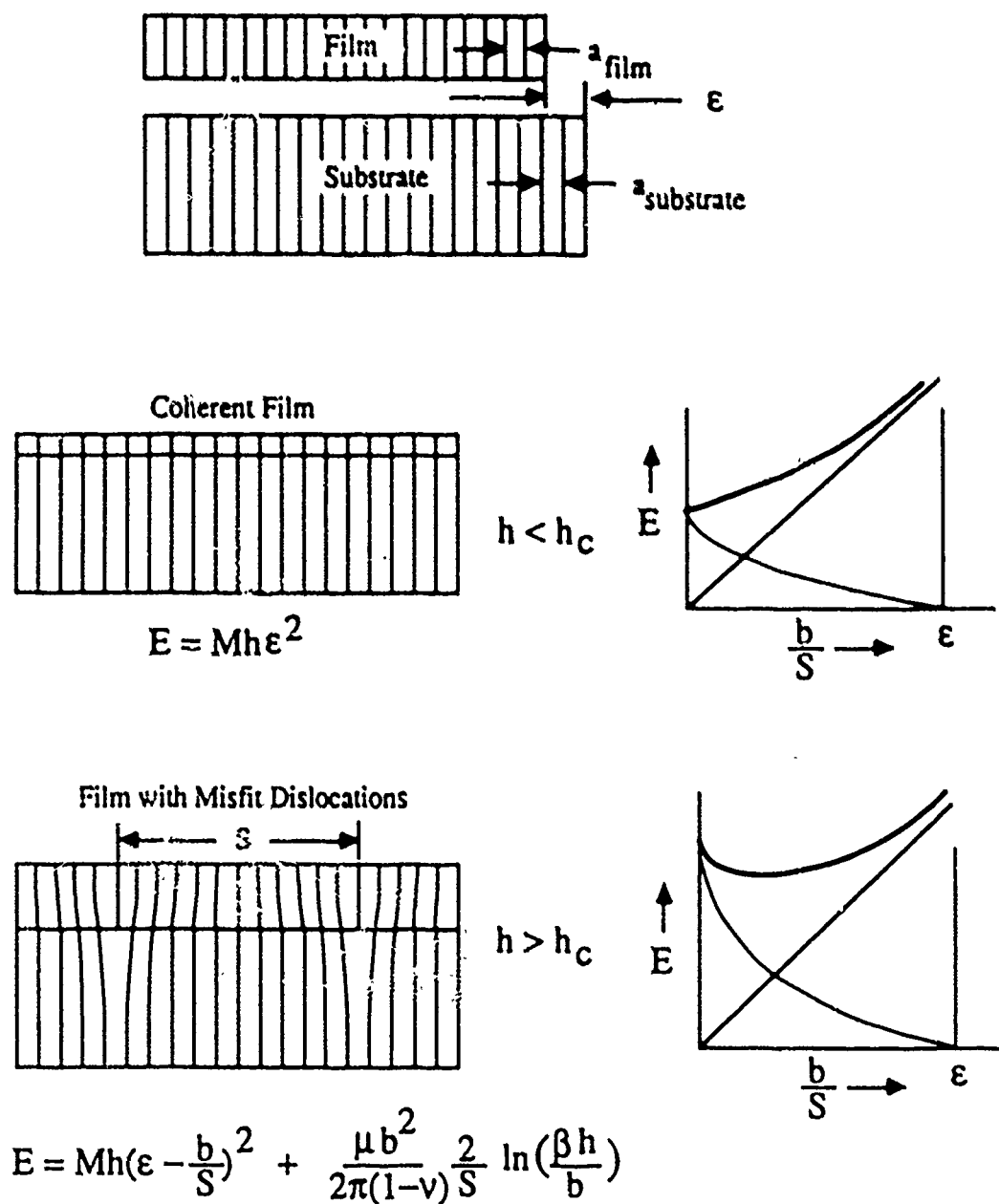
Figure 6.   Illustration of the Matthews' [5-7] equilibrium theory of misfit dislocation formation. Below a critical film thickness the equilibrium state of the film is dislocation free. Dislocations form in thicker films to reduce the free energy of the system.

for a very thin film is dislocation free can be shown by considering a slightly thicker film that contains misfit dislocations. The introduction of misfit dislocations partially accommodates the lattice misfit between the film and substrate and this allows the uniform strain in the film (and its associated energy) to be reduced. When dislocations with Burgers vectors b and spacing S are formed, the remaining homogeneous strain in the film is (ε - b/S) and the corresponding strain energy is reduced accordingly. While the strain energy associated with the homogeneous strain decreases as misfit dislocations are introduced, the energies of those dislocations increases the energy of the system. The dislocation energy (per unit area) is expressed approximately as

$$E_{\text{dislocations}} = \frac{\mu b^2}{4\pi(1-v)}\frac{2}{S}\ln\left(\frac{\beta h}{b}\right) \quad , \tag{15}$$

where μ is the shear modulus of thin film and substrate (here assumed to be the same), v is Poisson's ratio and β is a numerical constant of the order of unity. The factor 2/S in this expression represents the misfit dislocation length per unit film area, while the remaining terms give the energy per unit length of each misfit dislocation. The dislocation energy depends logarithmically on the film thicknesses because this dimension controls the outer cut-off radius for the elastic field of the dislocations. The total energy of thin films containing misfit dislocations is then found by adding these two terms. The result is

$$E = Mh\left(\varepsilon - \frac{b}{S}\right)^2 + \frac{\mu b^2}{4\pi(1-v)}\frac{2}{S}\ln\left(\frac{\beta h}{b}\right) \quad . \tag{16}$$

We note that the energy associated with the uniform strain (the first term in the RHS of eqn. (16)) varies linearly with the film thickness h, while the energy associated with the misfit dislocation varies only logarithmically with film thickness. As a consequence, only above a critical thickness, h_c, does the introduction of misfit dislocations lead to a decrease in the energy of the system. The equilibrium state of the system can be determined by finding the conditions for which the total energy (per unit area) reaches an absolute minimum with respect to the number of misfit dislocations per unit length, 1/S. This is obtained by setting the following derivative to zero:

$$\frac{\partial E}{\partial(\frac{1}{S})} = -2Mhb\left(\varepsilon - \frac{b}{S}\right) + \frac{\mu b^2}{2\pi(1-v)}\ln\left(\frac{\beta h}{b}\right) = 0 \tag{17}$$

The critical film thickness is then found by solving this equation for the special case of b/S = 0:

$$\frac{h_c}{\ln\left(\frac{\beta h_c}{b}\right)} = \frac{\mu b}{4\pi(1-\nu)Mc} \quad .$$

(18)

For h>h_c the equilibrium state includes misfit dislocations as shown in Fig. 6, but for h<h_c the lowest energy is achieved when no misfit dislocations are present. Thus, there is a critical thickness, below which fully coherent epitaxial films are thermodynamically stable.

Comparison with Experiment

Recent work on heteroepitaxial films of Si-Ge on Si suggests that these films remain dislocation free at thicknesses much greater than the critical thickness predicted by the equilibrium theory [8-10]. This result is shown in Fig. 7 for Si-Ge alloy films on Si substrates. Here, the data of Bean et al. [9] are compared with predictions of the equilibrium theory. For films containing 20 percent Ge or less, the actual film thicknesses at which misfit dislocations are observed are an order of magnitude greater than the critical thicknesses predicted by the equilibrium theory. These results have also been found recently by Gronet [11], as shown in Fig. 8. Similar results have been reported by Tsao et al. [10] for Ge-Si alloy films on Ge substrates. Some of their results are shown in Fig. 9. Although there was some debate about the evidence for this discrepancy [12-13], it now seems likely that the results cited here are genuine and that the cause of the discrepancy is related to the kinetics of nucleation, motion and multiplication of dislocations in epitaxial semiconductor films [10,14-16]. Below, we give a dislocation dynamics model for the formation of misfit dislocations in epitaxial thin films. The model provides a mechanistic understanding of misfit dislocation formation in terms of the kinetics of nucleation, motion and multiplication of misfit dislocations.

Following the work of Matthews [5-7], Freund [17] has shown that the most powerful way to understand the formation misfit dislocations in epitaxial structures is to consider the energetics associated with the incremental extension of a misfit dislocation, rather than considering the overall energy change associated with the formation of a periodic array of misfit dislocations. Because this approach is based on the mechanism of misfit dislocation formation, it is easy to see how the kinetics of dislocation motion should be included in the analysis.
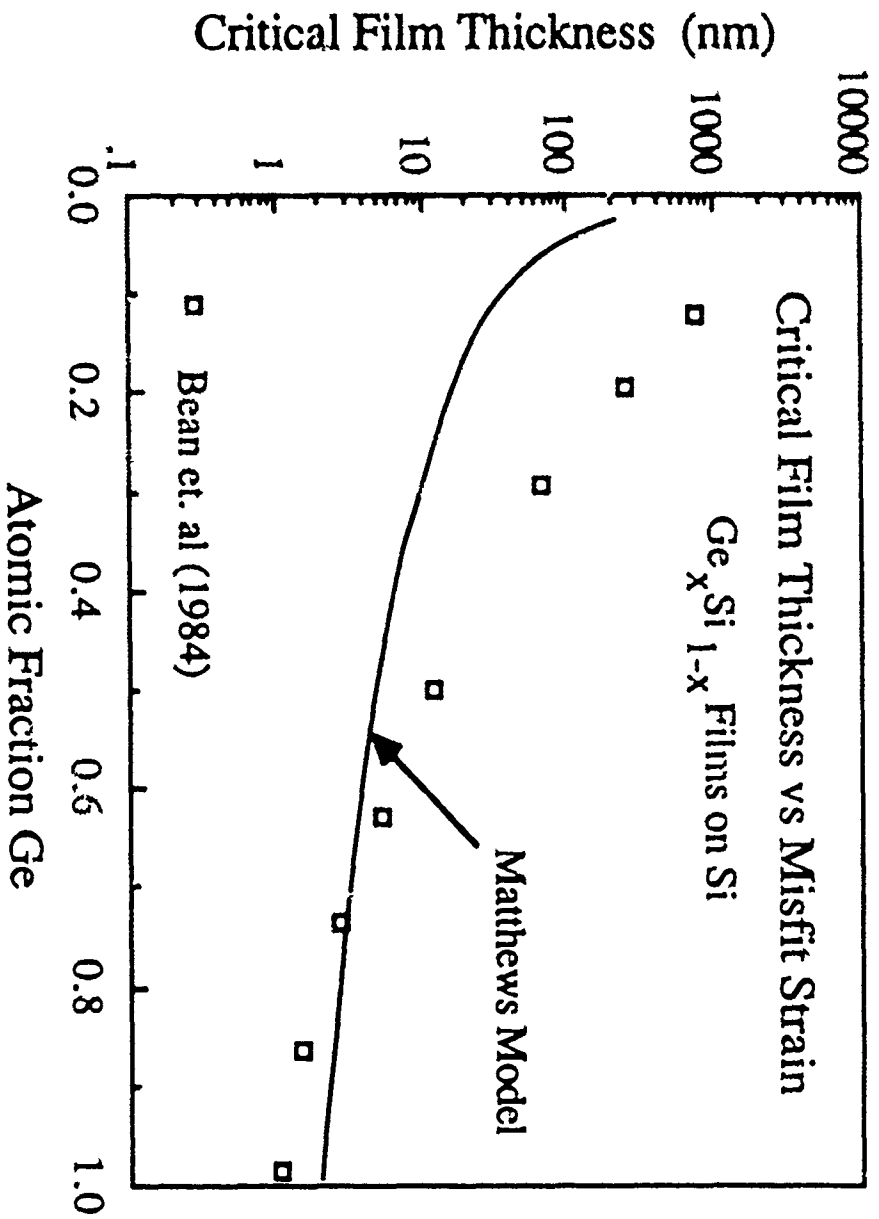
15

Critical Film Thickness vs Misfit Strain

Ge$_x$Si$_{1-x}$ Films on Si

Bean et. al (1984)

Matthews Model

Critical Film Thickness (nm)

Atomic Fraction Ge

Figure 7. Critical film thickness as a function of misfit strain for Ge$_x$Si$_{1-x}$ films on Si substrates. The strain in the film is determined by the composition. The experimental results of Bean et al. [9] are compared with Matthews' equilibrium theory [5-7].
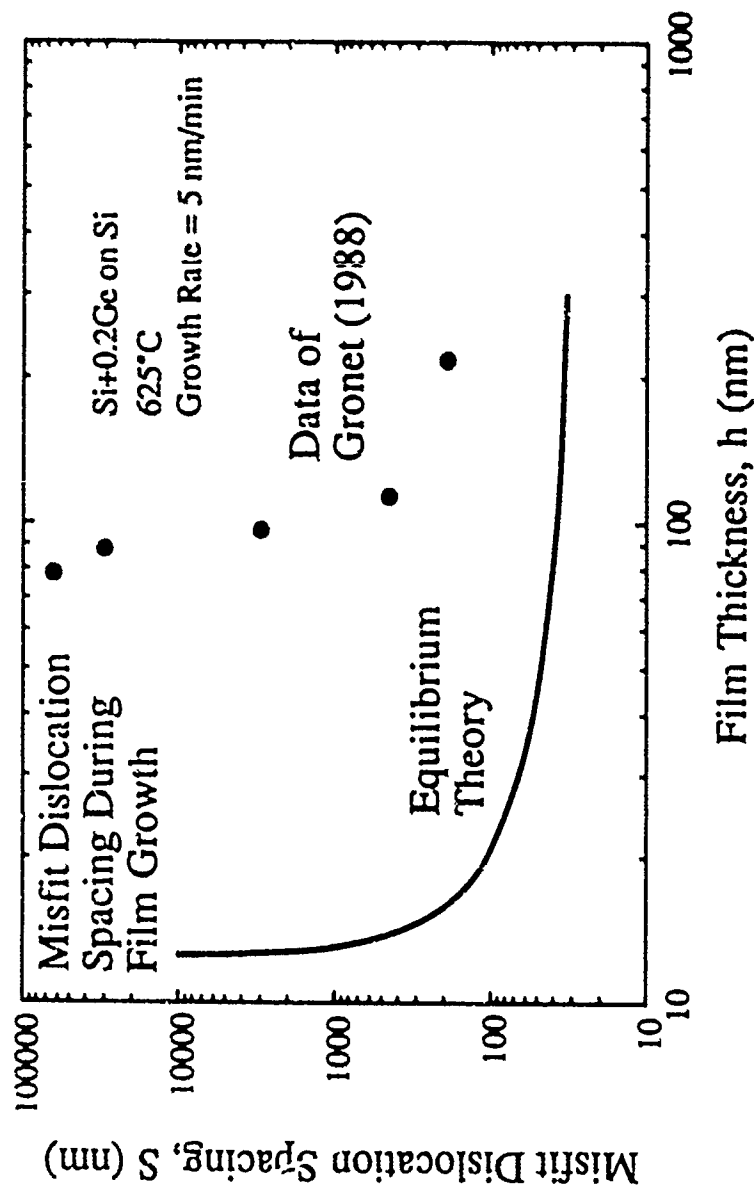
Figure 8.   Misfit dislocation spacing in Si-Ge films on Si substrates as a function of film thickness during film growth. The predictions of the equilibrium theory [5-7] are compared with measurements reported by Gronet [11]. The observed spacings are much greater than expected from the equilibrium theory.
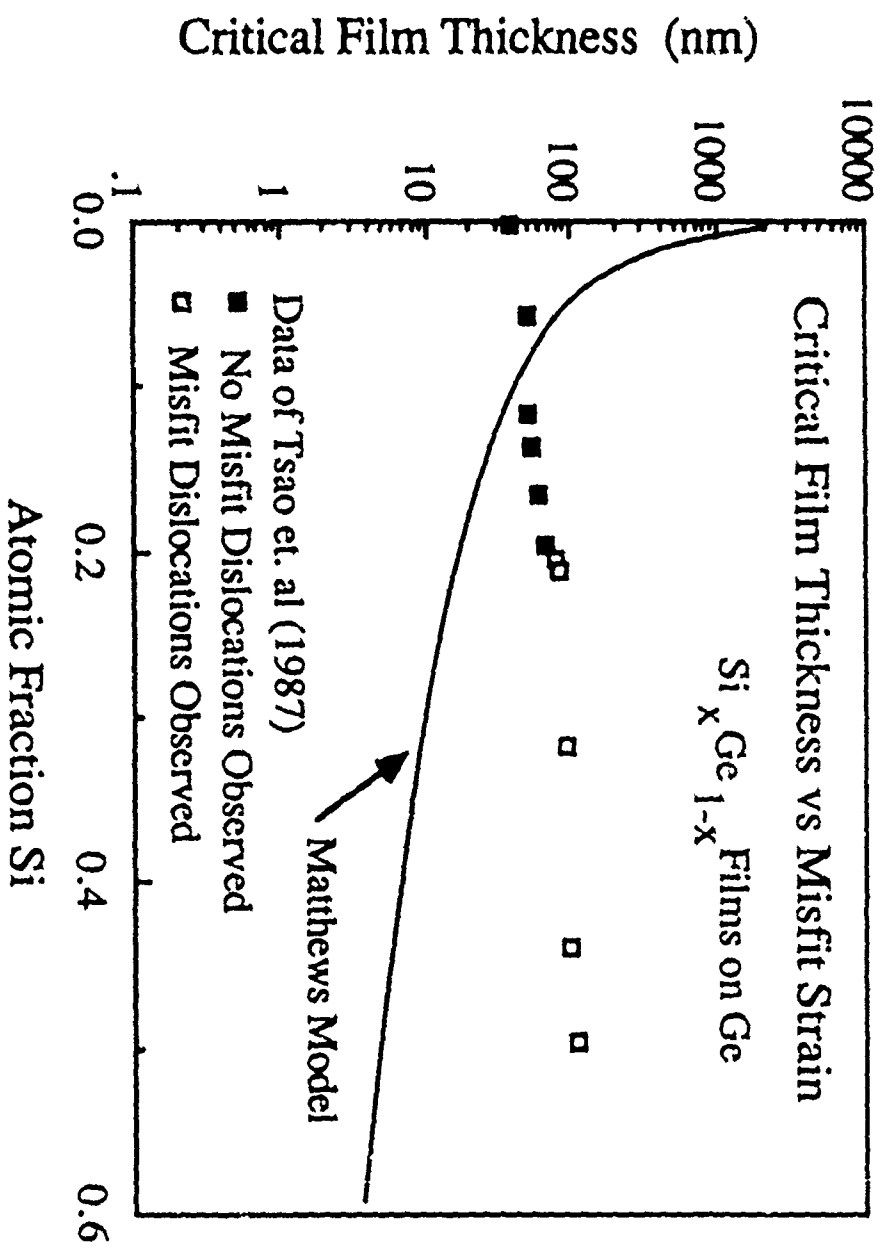
Critical Film Thickness vs Misfit Strain

$Si_x Ge_{1-x}$ Films on Ge

Data of Tsao et. al (1987)
- ■ No Misfit Dislocations Observed
- □ Misfit Dislocations Observed

Matthews Model

Atomic Fraction Si

Critical Film Thickness (nm)

Figure 9. Critical film thickness as a function of misfit strain (film composition) for Ge rich films on Ge substrates [10]. The composition (or strain) at which misfit dislocations are first observed is greater than predicted by the equilibrium theory [5-7].

Mechanisms of Misfit Dislocation Formation

Following the work of Freund [17], we envision a dislocation that extends from the free surface of a film to the film-substrate interface and deposits a misfit dislocation at the film/substrate interface as it moves. This is shown in Fig. 10. In this figure and throughout this section, we consider the case of FCC crystals with the (001) orientation and focus our attention on the motion of dislocations on the octahedral, {111}, planes. For films of this orientation, the angles defining the {111} slip plane normal and the Burgers vector are $\varphi = \cos^{-1}(1/\sqrt{3})$ and $\lambda = \cos^{-1}(1/\sqrt{2})$, respectively. Much of the recent experimental data has been obtained for Si-Ge alloy films grown onto pure Si substrates. Lattice mismatch causes the films to be in a state of biaxial compression, as shown in the figure.

Threading Dislocations

We may ask how the process shown in Fig. 10 gets started. If dislocations are present in the substrate on which the film is growing, then they grow naturally into the epitaxial film and reach the free surface of the film. These are sometimes called "threading" dislocations. The process by which threading dislocations can begin to form misfit dislocations at the film/substrate interface is shown schematically in Fig. 11. The biaxial stresses in the film exert forces on the threading dislocation and cause it to move in its slip plane. The portion of the dislocation that resides in the substrate remains stationary; the forces on it are much smaller and are of opposite sign. Thus the threading dislocation in the film bends over as it moves and eventually leaves a misfit dislocation in its wake. Continued movement of the threading dislocation extends the length of the misfit dislocation.

Misfit dislocations are also formed in epitaxial films grown onto dislocation-free substrates. In such cases, threading dislocations are not available to produce the misfit dislocations. Although the process is not well understood, misfit dislocations must be created by some form of dislocation nucleation, most probably at defects at the free surface of the growing film. Figure 12 shows how the nucleation of a dislocation half loop at the surface of the growing film eventually leads to the formation of a misfit dislocation at the film/substrate interface. The two ends of the half loop move in opposite directions and create additional misfit dislocation lengths as they move.
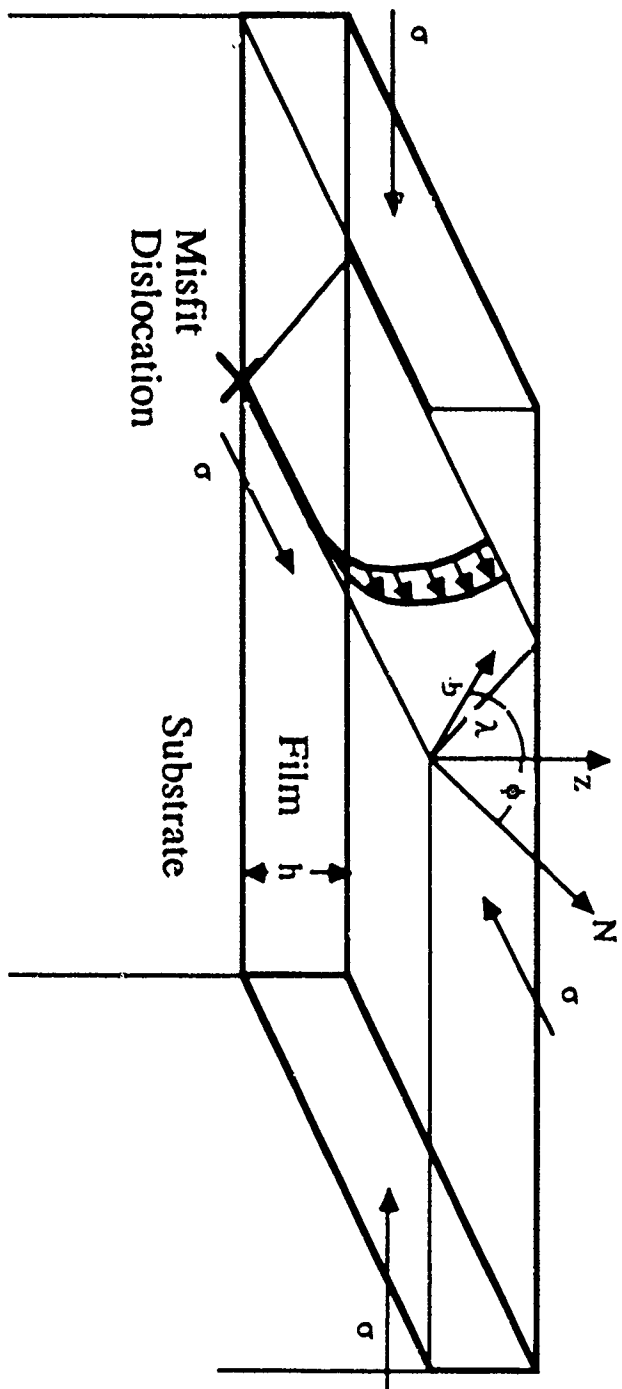
Figure 10. Dislocation motion in a thin film on a substrate leading to the deposition of a misfit dislocation at the interface between the film and the substrate.

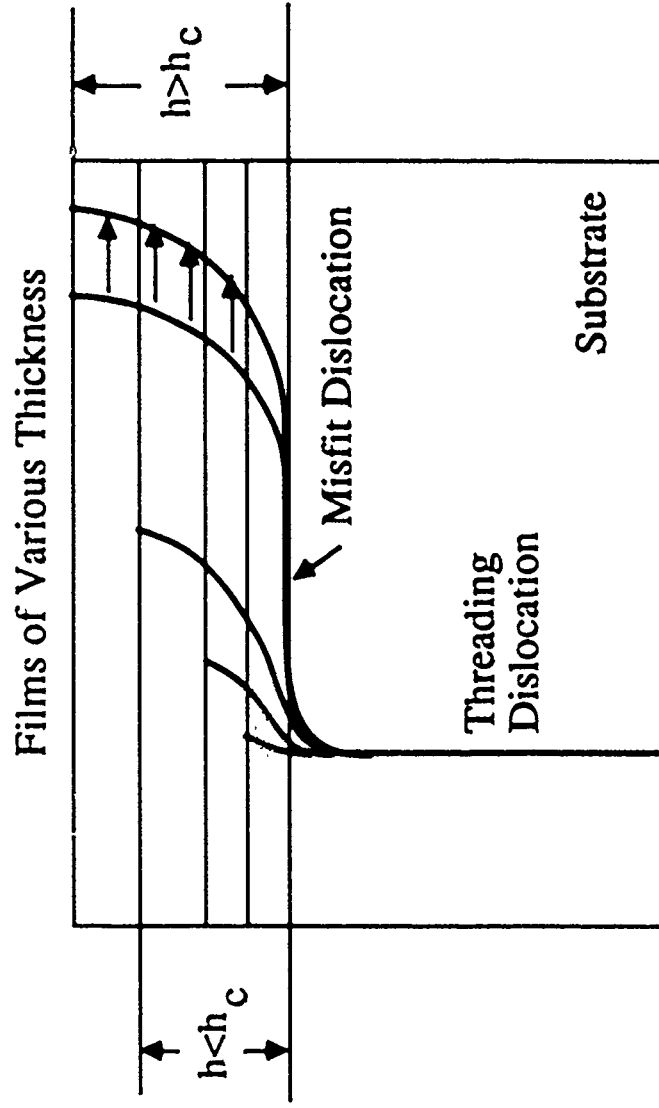# Bending of a Threading Dislocation During Film Growth

Films of Various Thickness

$h<h_c$

$h>h_c$

Misfit Dislocation

Threading Dislocation

Substrate

Figure 11. Bending of a threading dislocation in a strained film growing on a substrate. Above a critical thickness, the strain in the film is sufficient to cause the dislocation in the film to move and create a misfit dislocation.

Dislocation Nucleation at a Surface Source
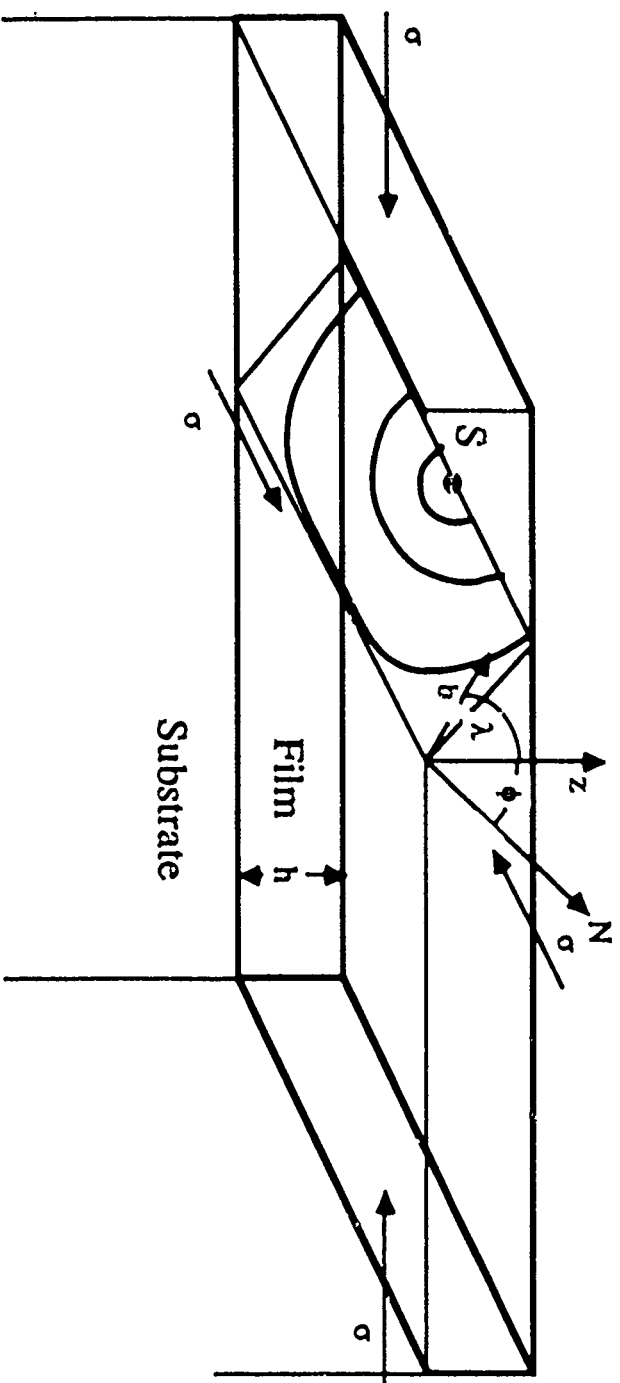
Figure 12. Dislocation nucleation at the surface of a strained film. The expanding half-loop eventually produces a misfit dislocation at the film-substrate interface.

Criterion for Misfit Dislocation Formation

Freund's model of misfit dislocation formation is based on the idea that the work done by the stress in the film when the dislocation moves a unit distance must provide enough energy to deposit a unit length of misfit dislocation. Following Freund, we call the work done by the stress in the film W_{layer} and call the work required to form a unit length of dislocation $W_{dislocation}$. For the dislocation geometry shown in Fig. 10 we can show that W_{layer} is

$$W_{layer} = \frac{\tau bh}{\sin \varphi} = \frac{\cos \lambda \cos \varphi}{\sin \varphi} \sigma bh = \frac{\sigma bh}{2} \quad . \tag{19}$$

For the FCC slip geometry under consideration, the Burgers vector makes an angle of 60° to the line of the misfit dislocation; thus, the misfit dislocation deposited at the film/substrate interface is a 60° mixed dislocation and $W_{dislocation}$ must be found for that case. We use the results of Freund [17] and Barnett [18] for edge and screw dislocations, respectively, to construct the solution for a 60° dislocation. Freund has shown that the energy of an edge dislocation lying at the interface between a thin film and semi-infinite substrate can be expressed as

$$W_{edge} = \frac{b_e^2}{4\pi(1-\nu)} \frac{2\mu_f \mu_s}{(\mu_f + \mu_s)} \ln\left(\frac{\beta_e h}{b_e}\right) \quad , \tag{20}$$

where (following the form given by Matthews) μ_f and μ_s are the shear moduli of the film and substrate, respectively, β_e is a numerical constant equal to 0.701, and the other terms have their usual meaning. The result for a screw dislocation was obtained recently by Barnett [18] and takes the similar form

$$W_{screw} = \frac{b_s^2}{4\pi} \frac{2\mu_f \mu_s}{(\mu_f + \mu_s)} \ln\left(\frac{\beta_s h}{b_s}\right) \quad , \tag{21}$$

where $\beta_s = 1.0$. We can obtain the result for a 60° mixed dislocation by noting that $b_e = (\sqrt{3}/2)b$ and $b_s = b/2$, and by adding the energies of the edge and screw components. After some rearrangement and using $\nu = 0.3$, we obtain

$$W_{60°dislocation} = 0.95 \frac{b^2}{4\pi(1-\nu)} \frac{2\mu_f \mu_s}{(\mu_f + \mu_s)} \ln\left(\frac{\beta h}{b}\right) \quad , \tag{22}$$

where $\beta = 0.755$.

We may now continue to develop a model for misfit dislocation formation. W_{layer} must be greater than $W_{60°dislocation}$ in order for the dislocations in the film to move and deposit misfit dislocations. On the basis of this, we may define an effective stress for dislocation motion in terms of the work done by the stress in the film less the work required to deposit the misfit dislocation:

$$\tau_{eff} b \frac{h}{\sin \varphi} = W_{layer} - W_{60°dislocation} \qquad (23)$$

This net driving force is positive only at film thicknesses greater than the critical thickness. Indeed, the critical film thickness discussed above can be found by setting the right side of this equation equal to zero.

$$W_{layer} = W_{60°dislocation} \qquad \frac{h_c}{\ln\left(\frac{\beta h_c}{b}\right)} = 1.9 \frac{b}{4\pi(1-v)\sigma} \frac{2\mu_f \mu_s}{(\mu_f + \mu_s)} \qquad (24)$$

where the biaxial stress σ has been used in place of the product of the biaxial modulus and misfit strain, $\sigma = M\varepsilon$. This result is a more exact form of eqn. (18) for this particular orientation. This relation may also be used to find the relaxed stress in a film containing an equilibrium number of misfit dislocations. For $h > h_c$ the result is

$$\sigma = \frac{1.9b}{4\pi(1-v)h} \frac{2\mu_f \mu_s}{(\mu_f + \mu_s)} \ln\left(\frac{\beta h}{b}\right) \qquad (25)$$

Kinetics of Dislocation Motion

The kinetics of misfit dislocation formation can be predicted using the effective stress defined by eqn. (23) and the dislocation mobility. Haasen and Alexander [19] have measured dislocation mobilities in Si and Ge, and others have made similar measurements for other semiconductors [20-21]. According to these studies, the dislocation velocities in semiconductors can be expressed as

18

$$v = B \exp\left(-\frac{U}{kT}\right)\left(\frac{\tau_{eff}}{\tau_0}\right)^{1.2} \quad , \quad \quad \quad . \quad \quad (26)$$

where τ_{eff} is the effective stress as defined above and U is the activation energy for dislocation motion. Because the effective stress is zero at the critical thickness, the rates of dislocation motion and misfit dislocation formation are both zero at $h=h_c$. The dislocations can move with finite velocities only in films that are greater than the critical thickness.

Because each moving dislocation deposits a misfit dislocation as it moves, the overall rate of misfit dislocation formation can be found by multiplying the velocity by the number of moving dislocations per unit area. If the misfit dislocation density, defined as the line length of misfit dislocations per unit area of film, is ρ_{mf}, it follows that the rate of production of misfit dislocations can be expressed as

$$\frac{d\rho_{mf}}{dt} = Nv \quad , \quad \quad \quad (27)$$

where N is the moving dislocation density. Thus, the rate of formation of misfit dislocations depends on the density of mobile dislocations that are already there. If the substrate is dislocation-free and no dislocations are nucleated in the epitaxial film during growth, then misfit dislocations cannot form.

Constant Threading Dislocation Density The simplest kinetic model is one in which a fixed number of threading dislocations is assumed to exist in the substrate prior to film growth. Such dislocations would extend into the growing film and would be subjected to the stresses there. As the film thickness increases, above the critical film thickness, these dislocations bend over and begin to glide parallel to the film-substrate interface, creating misfit dislocations as they move. This process is shown in Fig. 11. If the density of threading dislocations is constant (no multiplication of existing dislocations or nucleation of new ones), the rate of increase of misfit dislocation line length is governed entirely by the dislocation velocity. If we define the average spacing between misfit dislocations, S, as

$$S = \frac{2}{\rho_{mf}} \quad , \quad \quad \quad (28)$$

then the spacing between misfit dislocations will decrease continuously during the course of growth once the critical film thickness has been exceeded. Figure 13 shows how the misfit dislocation spacing in epitaxial films of Si+0.2Ge on Si is expected to change during film growth. Various constant mobile dislocation densities were assumed in the calculations. The film was assumed to be growing at a rate of 5 nm per minute at a temperature of 625°C and the velocities of the dislocations were calculated using eqn. (26) with U=2.2eV, τ_0= 9.8 MPa and B= 7.33x10^4 m/s. The curve indicating the largest misfit spacings corresponds to the lowest mobile dislocation density. The lowest curve corresponds to the equilibrium spacing between misfit dislocations. For all curves, the misfit spacings tend to infinity at a film thickness of about 13.5 nm. This corresponds to the critical film thickness for this system. One notes that the curve corresponding to the largest dislocation density coincides with the equilibrium theory at large film thicknesses, as expected.

Multiplication of Dislocations   The above treatment is unrealistic because it does not allow for either dislocation multiplication or nucleation of new dislocations in the course of growth. Several mechanisms of dislocation multiplication can be envisioned. One of these, suggested first by Hagen and Strunk [22], has been directly observed by Rajan and Denhoff [23] using weak beam TEM techniques. According to this mechanism, two crossing misfit dislocations with the same Burgers vectors, of the kind shown in Fig. 14, can annihilate locally and produce two new mobile dislocations, which can then glide and produce additional lengths of misfit dislocation. This is but one of many mechanisms that are possible. All such mechanisms, however, can be described using the concept of a breeding factor. We define the breeding factor, δ, as the number of new dislocations produced per unit length of glide motion by each moving dislocation. Then the multiplication of dislocations can be described by

$$\frac{dN}{dt} = N\delta v \qquad . \tag{29}$$

The effect of dislocation multiplication on the misfit dislocation spacing during the course of film growth is shown in Fig. 15. Here we have assumed a very low initial dislocation density of N=0.1 cm^{-2} and have considered constant breeding factors of 5, 10 and 20 dislocations/mm. We note that although some dislocation motion occurs as soon as the critical film thickness is exceeded, the misfit dislocation spacing remains very large until the film has reached a thickness of about 100 nm. The predictions are in qualitative agreement with the results of Gronet [11] shown in Fig. 8.
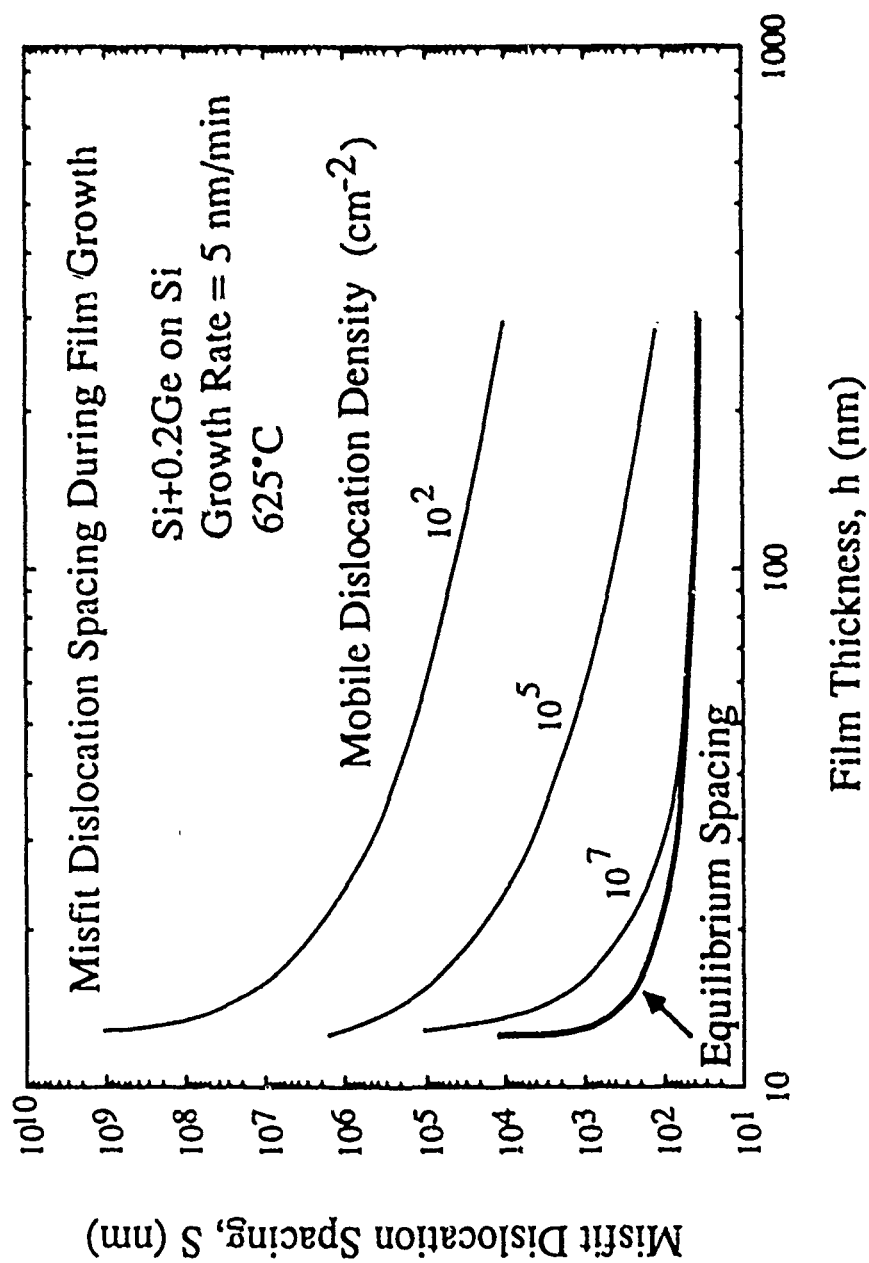
20

Figure 13. Misfit dislocation spacing as a function of film thickness during growth of a Si+0.2Ge film on a Si substrate under the conditions shown. Calculations are shown for three different threading dislocation densities. The predictions of the equilibrium theory are also shown.
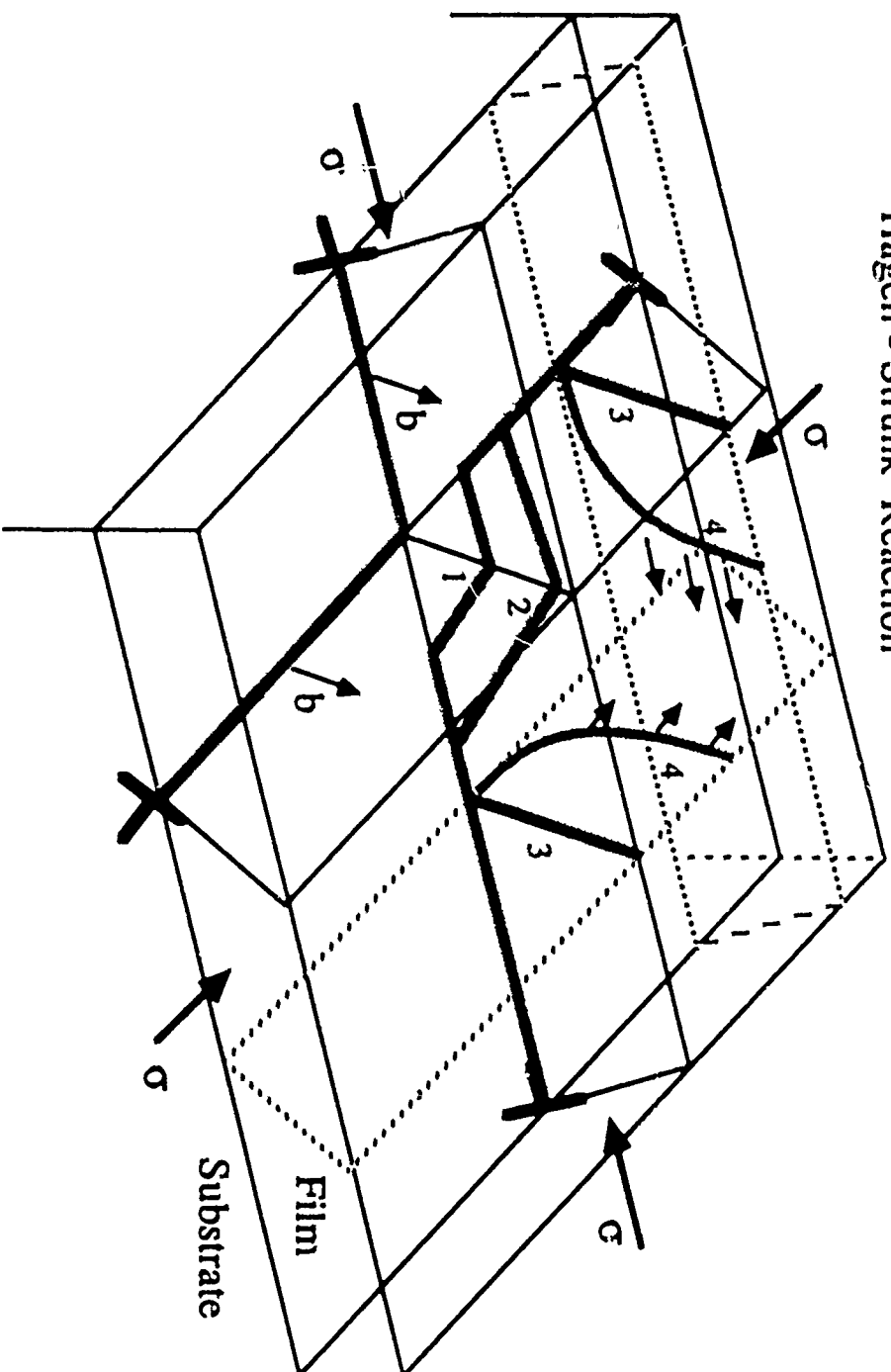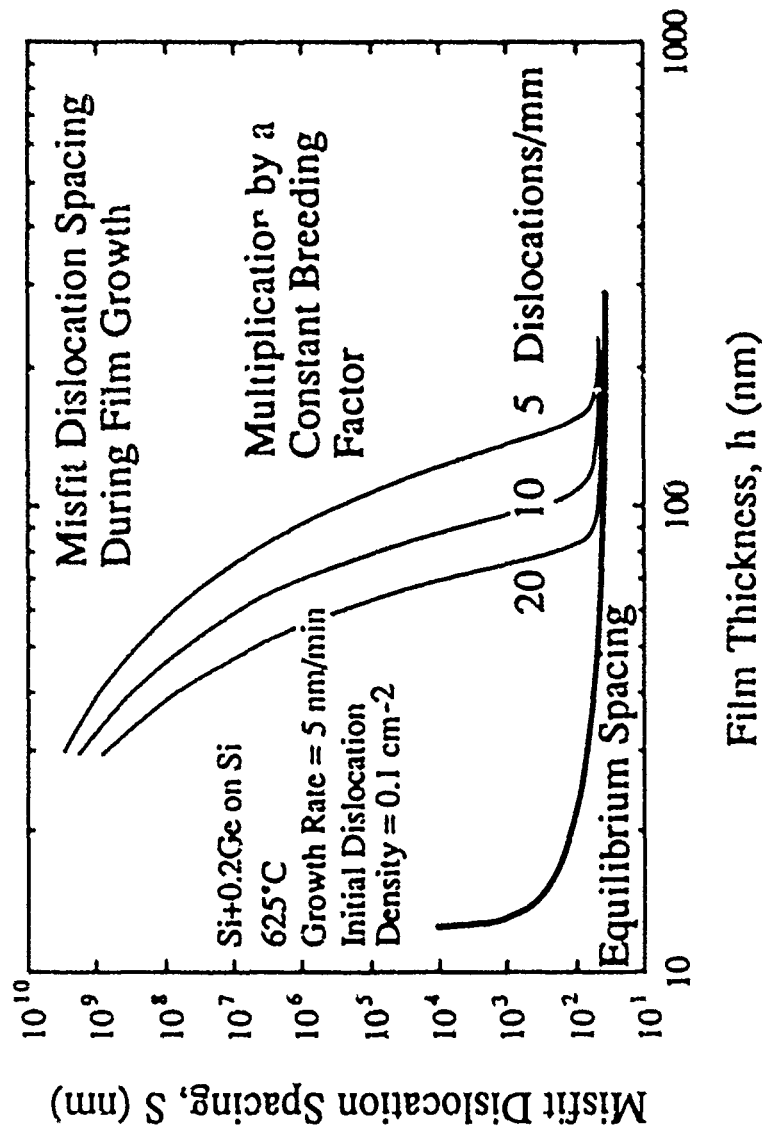
Hagen - Strunk Reaction

Figure 14. Illustration of the Hagen-Strunk [22] mechanism of dislocation multiplication. Crossing dislocations with the same Burgers vectors can react to form two new mobile dislocations as shown.

Figure 15. Misfit dislocation spacing as a function of film thickness during growth of a Si+0.2Ge film on a Si substrate under the conditions shown. Calculations are shown for three different rates of dislocation multiplication. The predictions of the equilibrium theory are also shown.

According to the Hagen-Strunk multiplication mechanism discussed above, the breeding factor should be proportional to the number of Hagen-Strunk type intersections produced per unit length of dislocation travel. This may be expressed as

$$\delta_{HS} = \frac{P_{HS}}{2S} \quad , \tag{30}$$

where S is the current spacing between misfit dislocations and p_{HS} is breeding efficiency, or the probability that any particular Hagen-Strunk intersection actually results in a new pair of moving dislocations. The effect of this kind of multiplication on the spacing of misfit dislocations is shown in Fig. 16. Here we consider the case of a film with an initial dislocation density of $N_0=1000$ cm^{-2}. We show the effect of Hagen-Strunk multiplication by calculating the evolution of the misfit dislocation spacing for various breeding efficiencies. The calculation involves coupling eqn. (27) with eqn. (28), and using eqns. (29) and (30) to describe the multiplication process. We see, as expected, that multiplication can have a profound effect on the evolution of the misfit dislocation spacing. Generally, increasing the breeding factor causes the misfit dislocations to form more abruptly during film growth. Indeed, the curves of Figs. 15 and 16 give the impression that the critical thickness for misfit dislocation formation is much greater than the true critical thickness of 13.5 nm. This helps to explain why misfit dislocations are often not observed in films that have been grown well beyond the critical thickness.

Nucleation of Dislocation Half-Loops   New dislocations can also be formed by nucleation in the growing film, perhaps at the growing surface. Such nucleation must occur in those cases where misfit dislocations are formed in epitaxial films grown on dislocation-free substrates. For the present analysis we let $(dN/dt)_0$ represent the rate of nucleation of new dislocations. Then the total rate of formation of new dislocations can be expressed as

$$\frac{dN}{dt} = \left(\frac{dN}{dt}\right)_0 + N\delta v \quad . \tag{31}$$

When new moving dislocations are formed in the course of growth, the rate of formation of misfit dislocations is described by coupling eqn. (31) with eqn. (27). The effect of dislocation nucleation on the evolution of the misfit dislocation spacing is shown in Fig. 17 for various rates of nucleation. We see that increasing the rate of nucleation causes the misfit dislocation spacing to more quickly approach the equilibrium spacing. Nucleation of dislocations during the course of
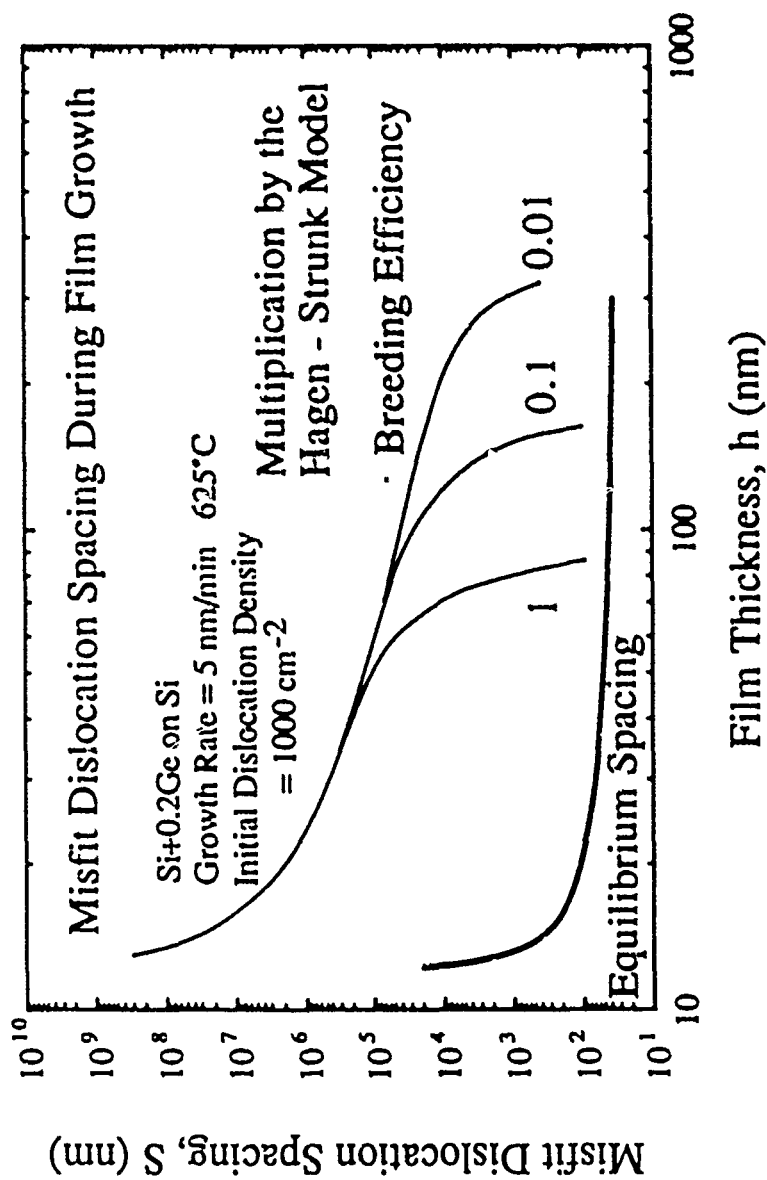
Figure 16. Misfit dislocation spacing as a function of film thickness during growth of a Si+0.2Ge film on a Si substrate under the conditions shown. Calculations are shown for the Hagen-Strunk multiplication mechanism with three different efficiencies. The predictions of the equilibrium theory are also shown.
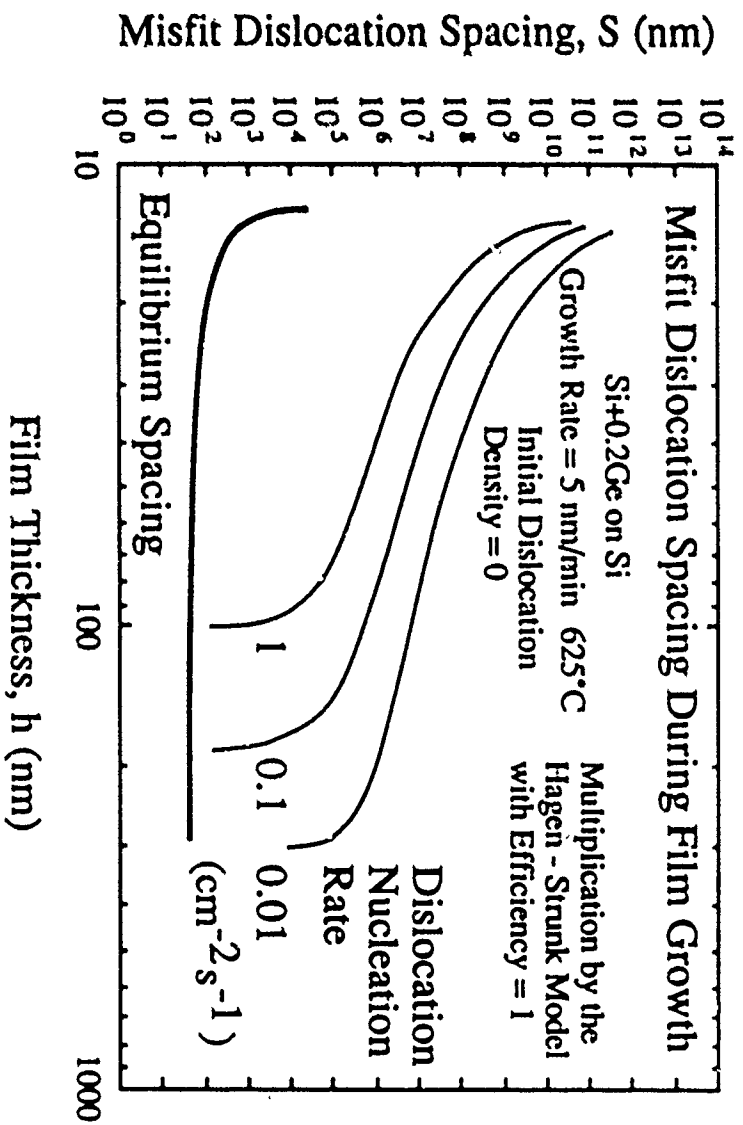
Misfit Dislocation Spacing During Film Growth

Misfit Dislocation Spacing, S (nm)

Si+0.2Ge on Si

Growth Rate = 5 nm/min 625°C

Initial Dislocation
Density = 0

Multiplication by the
Hagen - Strunk Model
with Efficiency = 1

Equilibrium Spacing

1   0.1   0.01
($cm^{-2}s^{-1}$)

Dislocation
Nucleation
Rate

y-axis: 10^{14}, 10^{13}, 10^{12}, 10^{11}, 10^{10}, 10^{9}, 10^{8}, 10^{7}, 10^{6}, 10^{5}, 10^{4}, 10^{3}, 10^{2}, 10^{1}, 10^{0}

Film Thickness, h (nm)

x-axis: 10, 100, 1000

Figure 17. Misfit dislocation spacing as a function of film thickness during growth of a Si+0.2Ge film on a Si substrate under the conditions shown. Calculations are shown for three different rates of dislocation nucleation. The predictions of the equilibrium theory are also shown.

film growth makes the misfit dislocations form even more abruptly than when only multiplication effects are included in the analysis.

We note that stress relaxation during the course of film growth has been taken into account in all of the above calculations. The creation of misfit dislocations is equivalent to plastic deformation in the film, and relieves the elastic strain and stress in the film. The corresponding evolution of the stress is described by

$$\frac{d\sigma}{dt} = -\frac{MNb}{2}v \quad , \tag{32}$$

where M is the biaxial elastic modulus of the film.

V.   Biaxial Strengths of Thin Films on Substrates

We now turn our attention to the strengths of thin films on substrates. We will see that the basic processes of misfit dislocation formation discussed above can also be used to understand plastic deformation of thin films on substrates. In addition, the relations we have developed for misfit dislocations can be used to describe the high strengths of these thin film materials.

Yielding and plastic flow in a thin film are, of course, driven by the stresses present in the film. Thus, any experimental study of biaxial strength must start with a measurement of the stress in the film. As in the case of bulk materials, the onset of plastic flow, or yielding, occurs when increments in the stress lead to non-elastic strain increments in the film. But first we must address the problem of measuring stress in the film.

Techniques for Measuring Thin Film Stresses

The experimental techniques for measuring stresses in thin films on substrates fall into two general classes: (1) those based on direct measurements of the elastic strains in the films using x-rays and (2) those based on the associated curvature or deflection of the substrate. The direct x-ray techniques are most informative because they permit a measurement of all of the components of stress in the film. They can, in principle, be used to detect spatial variations of stress within the film, either from grain to grain or from one point in the film to another. They can also be used to find the stresses in patterned films with irregular geometries. However, because these techniques are based on diffraction, they are limited to crystalline films; they cannot be used to find the stresses in non-crystalline materials such as passivation glasses or amorphous oxides. For these cases the stresses must be determined by measuring the curvature or deflection of the substrate. Even for crystalline films, the substrate curvature technique is often preferable because it is more convenient to use and easier to apply to special conditions. Such conditions include in-situ heating or cooling or stress measurements during the course of film growth.

We saw in Section III above that the biaxial stress in a thin film on a much thicker substrate is directly proportional to the associated curvature of the substrate, provided the substrate deforms elastically. For the usual case of a substrate that is elastically isotropic in the plane of the film, the expression for the biaxial stress in the film is

23

$$\sigma_f = M_s \frac{h_s^2}{6h_f} K = M_s \frac{h_s^2}{6h_f} \frac{1}{R} \quad , \tag{33}$$

where $K = 1/R$ is the curvature of the substrate, M_s is the biaxial modulus of the substrate and h_f and h_s are the thicknesses of the film and substrate, respectively. We see that the film stress does not depend on the properties or behavior of the film. This relation is valid for both elastic and plastic deformation in the film. In most cases the bare substrate is not perfectly flat so that the curvature K in eqn. (33) must be replaced by the change in curvature associated with the presence of the film. Thus the stress in a film is found by measuring the curvature of the substrate both before and after the film is deposited, or equivalently, before and after the film is removed from the substrate.

X-Ray Diffraction

A variety of techniques can be used to measure the curvature changes associated with thin film stresses. For the case of single crystal substrates, the curvature of the substrate produces a curvature of the crystal lattice that can be detected by x-ray diffraction. An advantage of this technique is that for a single film the technique is absolute. In such a case it is not necessary to measure the curvature before the film is deposited or after the film is removed from the substrate, because it is known that for most good substrates, such as silicon, the crystal lattice would have no curvature in either case.

Optical Interferometry

Optical interferometry represents another method for measuring changes in substrate curvature. A simple count of interference fringes permits a determination of the curvature. This technique is preferable if a more general measurement facility is not available and if the induced curvature is quite large. The principle of interferometry could be used to create a rapid, quantitative stress measuring technique comparable with the laser scanning technique described below, but as yet such instruments are not available.

Laser Scanning

The laser scanning technique is the most popular technique for measuring curvature changes associated with thin film stresses. The principle of the technique is quite simple. A beam of laser

24

light reflects off the surface of a curved substrate at an angle θ that depends on the orientation of the surface. Upon moving (or scanning) the laser beam to a new position, the light reflects at a different angle if the substrate is curved. A position sensitive photodetector can be used to detect the chang in angle of the reflected laser beam. Flinn [24] has recently developed a laser scanning device that makes use of a rotating mirror to scan the laser beam over the substrate. A schematic diagram of his instrument is shown in Fig. 18. A lens is used to cause the rotating laser beam to be perpendicular to the substrate at all points in the scan and the same lens causes the reflected beams to converge to one point on the position sensitive photodetector if the substrate is perfectly flat. When the substrate is curved, the reflected beam moves to different positions on the photodetector as the laser beam is scanned across the substrate. The linear motion detected by the photodetector can be converted to changes in angle as a function of position on the substrate and this, in turn, can be used to find the curvature of the substrate.

The laser scanning device is very sensitive; it is capable of detecting the bending of a flat wafer to a radius of about 10 km. For typical film dimensions, this corresponds to a film stress of about 0.2 MPa. This sensitivity is required because the film, being so much thinner than the substrate, bends the substrate only very slightly. As noted above, bare substrates are usually not perfectly flat. They often take the form of a potato chip, although typical substrates are, of course, much flatter than potato chips. For such a shape the curvature varies from point to point on the substrate and depends on the path of the scan. The shape of the substrate along a particular laser scan path must be found before the film is deposited, or after the film is removed, in order to obtain the change in curvature associated with the presence of the film. Even though the curvature of a substrate is not constant with position along the scan, the change in curvature associated with the film is often quite constant, provided that the thickness of the film is constant and the deposition conditions are the same at all points on the substrate. In such cases the subtraction technique described above gives a single value for the change in curvature and this in turn leads to a single value for the average biaxial stress in the film.

Throughout this discussion we have assumed that the biaxial stress in the film does not vary through the thickness of the film. If the stress varies through the film thickness, the substrate curvature technique gives only the average value of the stress. Doerner and Brennan [25] have recently used a grazing incidence x-ray scattering technique to measure the biaxial stresses in films of aluminum on silicon as a function of distance from the film-substrate interface. Some of their results are shown in Fig. 19. The samples were initially heated to 450°C and allowed to relax to a stress-free state before cooling to room temperature where the measurements were made. On cooling to room temperature, much of the differential thermal strain is accommodated by plastic

25

Laser Scanning Technique for Measuring Substrate Curvature

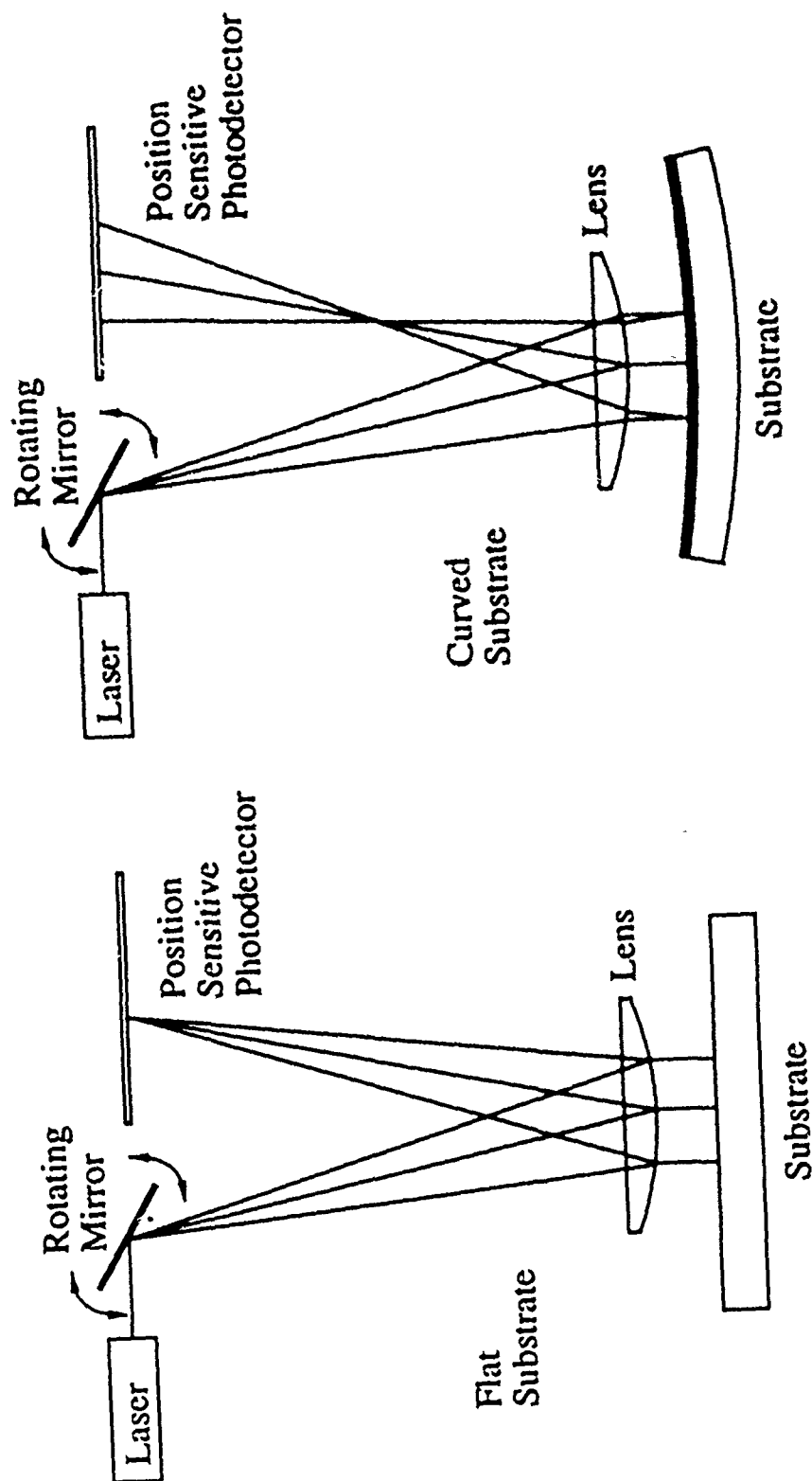## Motion of Reflected Beam Indicates Substrate Curvature

Figure 18. Schematic diagram of a laser scanning instrument used to measure substrate curvature and associated film stress [24].

Strain Distribution in Al Films on Si Substrates

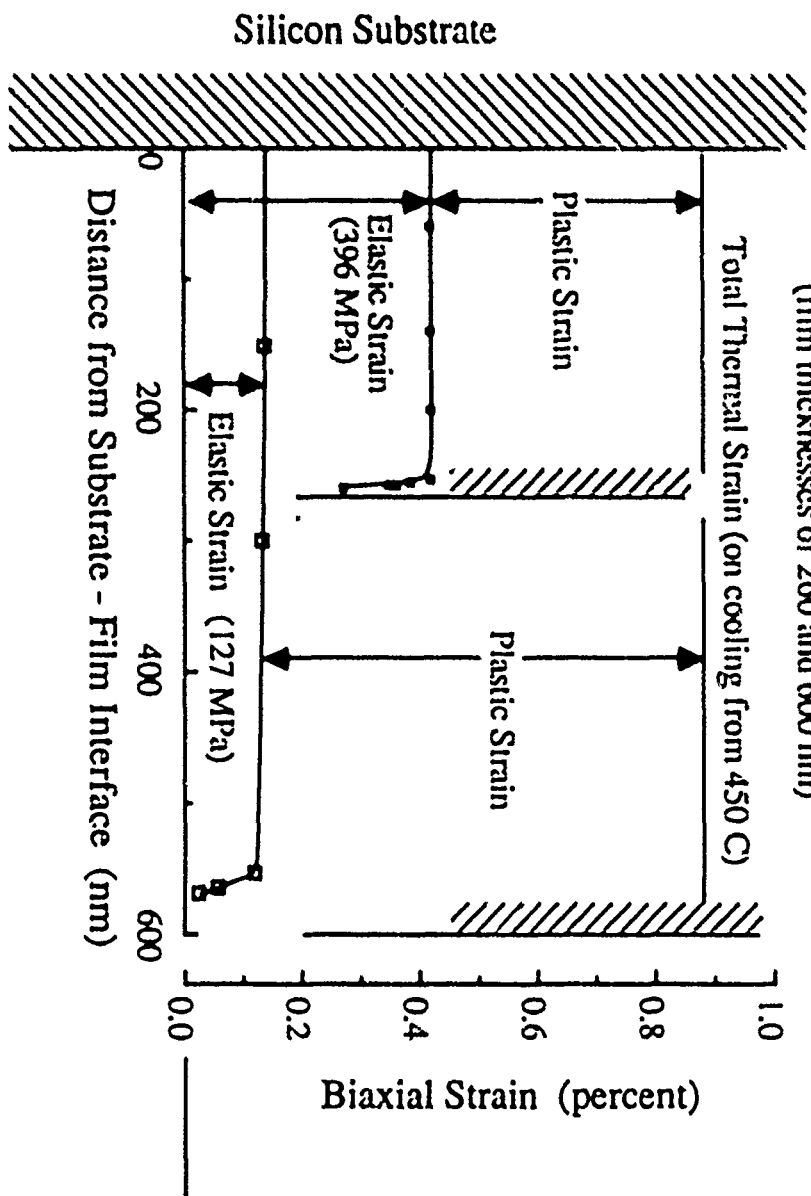GIXS Technique (Doerner and Brennan, 1987)
(film thicknesses of 260 and 600 nm)

Figure 19. Elastic strain distribution in Al films on Si substrates obtained using the GIXS technique [25]. The strains (and stresses) are uniform over most of the film thickness.

deformation, leaving a small residual elastic strain in the film. The x-ray technique gives the remaining elastic strain (or stress) as a function of distance from the film-substrate interface. Except for a sharp variation very near the surface of the film, the elastic strain (and stress) remains relatively constant through the thickness of the film. Thus, for this case, the substrate curvature technique can be used to determine the local biaxial stress through most of the thickness of the film. This assumption is probably not valid for films in their as-deposited state because their structure (and stress) may vary significantly through the thickness of the film.

Measurement of Elastic and Plastic Deformation in Thin Films on Substrates

In order to study the deformation properties of a thin film on a substrate, it is necessary to be able to vary the stress (or strain) and measure the corresponding changes in strain (or stress) in the film. Because the film is usually quite thin and because it is attached to the substrate, ordinary bulk mechanical testing methods cannot be used. However, we can use thermal expansion and the substrate curvature method to create a thin film mechanical testing technique. We can make use of differences in thermal expansion coefficients between the film and the substrate to impose varying biaxial strains in the film through heating and cooling. The curvature technique can then be used to measure any corresponding changes in film stress.

The results of such a testing procedure are shown in Fig. 20 for the case of a polycrystalline film of aluminum on an oxidized silicon wafer. Here the substrate curvature method was used to determine the biaxial stress in the film as a function of temperature during one of several heating and cooling cycles. The film had been annealed at 450°C prior to conducting the experiment in an effort to stabilize the grain structure. This annealing step eliminates the fine grained microstructure found in the as-deposited state and removes special features of the stress-temperature history that occur only during the first heating cycle. It also leaves the film in a state of biaxial tension at room temperature. As shown in Fig. 20, the biaxial tensile stress in the film at room temperature is about 280 MPa at the beginning of the second heating cycle. As the film-substrate composite is heated, the greater thermal expansion of the aluminum film first relaxes the tensile stress in the film and later causes the film to go into a state of compression. During the initial unloading portion of the curve, the differences in thermal expansion between the film and substrate are accommodated by elastic deformation of the film. The slope of the curve in this regime is simply the product of the difference in thermal expansion coefficients and the biaxial elastic modulus of the film, as shown in the figure. Eventually biaxial yielding of the film in compression is observed as a deviation from the thermo-elastic loading line. For a more precise definition, the biaxial yield stress would have to be determined using a suitable plastic offset, such as 0.2%. With continued
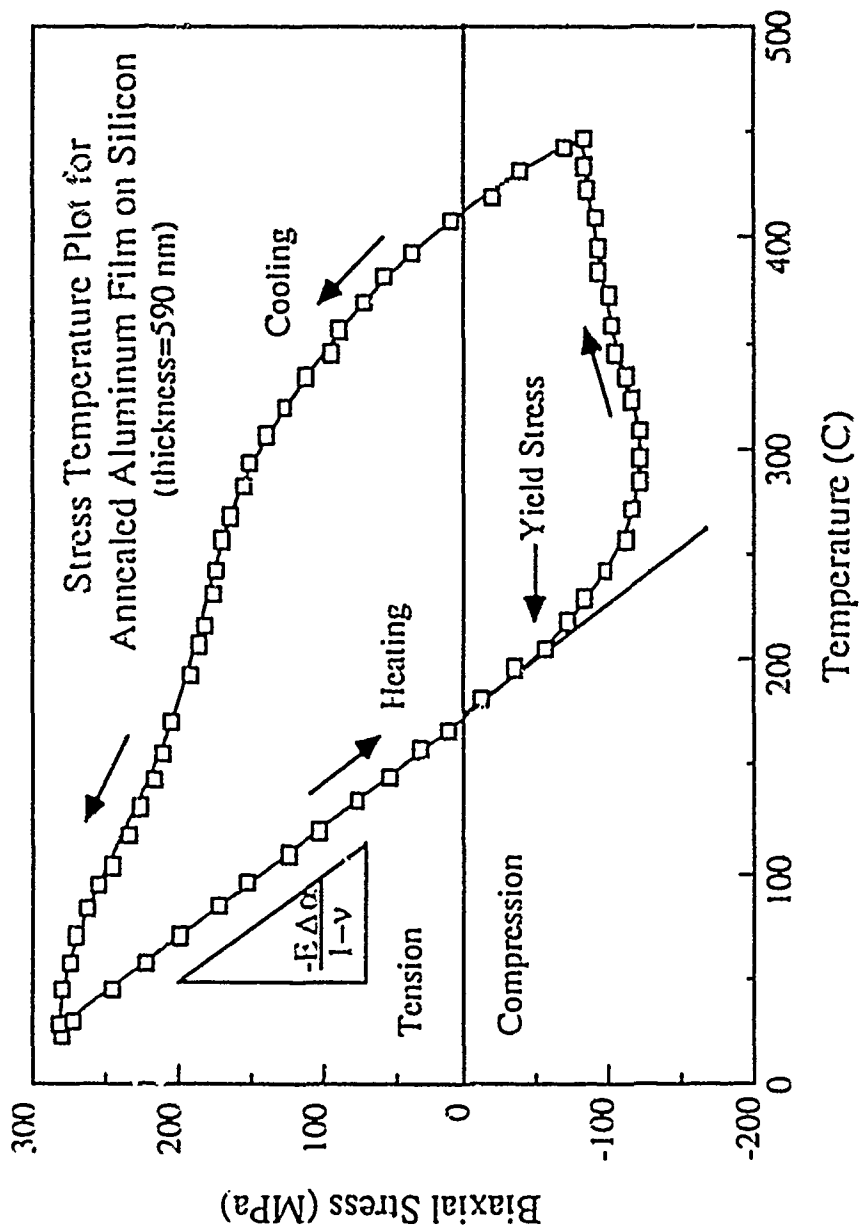
26

Figure 20. Stress - temperature plot for an Al film on a Si substrate. Elastic and plastic deformation occur in the film during heating and cooling.

heating after yielding, the compressive stress rises only slowly and eventually passes through a maximum. In this part of the experiment the changing thermal strain is accommodated by compressive plastic deformation in the film. The decreasing compressive stress with increasing temperature is an indication of the expected temperature dependence of the flow stress. On cooling from 450°C, the compressive stress is first relaxed to zero as the film thermally contracts more than the substrate. On further cooling, a tensile stress develops in the film and this causes plastic deformation in tension to occur. The stress-temperature curve does not go through a maximum on cooling because the flow strength increases with decreasing temperature. The curve in this part of the experiment corresponds to the flow stress of aluminum as a function of temperature. We note that the stress in the film after the thermal cycle is complete is about the same as the stress at the start of the cycle. This is not unique to this particular experiment; it is commonly observed for metal films on silicon. In fact, repeated heating and cooling cycles such as the one described here result in stress-temperature plots that are almost identical to the one shown in Fig. 20.

The stress-temperature history for an alloy of Al+2%Cu that had also been annealed once prior to testing is shown in Fig. 21 [26]. The main features of the curve are the same as those for pure aluminum. In this case, however, the microstructure changes with temperature as the copper dissolves on heating and precipitates from solution on cooling. The effect of precipitation from solution on cooling is indicated by the rapid strengthening that occurs near 150°C. Below that temperature the film does not deform plastically, indicating that the yield strength exceeds the actual stress in the film.

This technique for studying the plastic deformation properties of thin films on substrates permits one to study both yielding and plastic flow. Even time-dependent plastic flow can be studied by holding the temperature constant and measuring the kinetics of stress relaxation. However, it should be noted that because the strains are thermally imposed, generally it is not possible to conduct an isothermal stress-strain deformation experiment. Thus, the stress-temperature diagrams shown in Figs. 20 and 21 represent non-isothermal mechanical hysteresis curves.

Microstructural Changes in Thin Films

In the examples cited above, the stress-temperature relations are dominated by the effects of elastic and plastic deformation. Any microstructural changes that occur during heating and cooling are assumed to influence the resulting stress-temperature curve only by their effects on the elastic and plastic properties. When major microstructural changes occur, this assumption is no longer
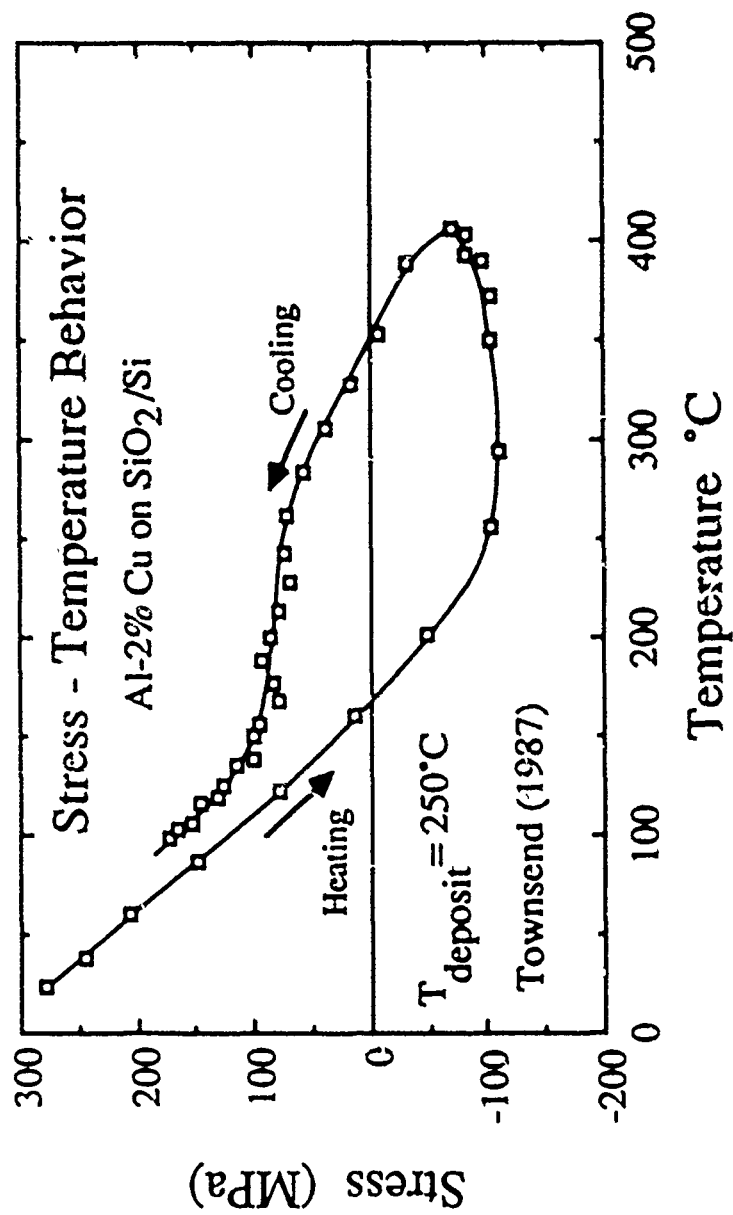
27

Figure 21. Stress - temperature plot for an Al+2%Cu film on a Si substrate [26].

valid and the dilatational effects of microstructural changes must be included in the interpretation of the stress-temperature curve.

As noted above, the stress-temperature history for the first heating of an as-deposited film is usually quite different from the curves discussed above. An example is shown in Fig. 22 [27]. This film is in biaxial tension in the as-deposited state. Upon heating, the greater thermal expansion of the film compared to the substrate causes the stress to become compressive. With continued heating the compressive stress in the film reaches a very large value before decreasing rapidly in the temperature range of 225-275°C. This stress relaxation is not caused by plastic deformation in the film. It is the result of the densification of the film that accompanies grain growth. In the as-deposited state, the grain size in the film is extremely fine. This as-deposited grain structure is unstable; the grains begin to grow at about 225°C and continue to grow during heating until they are comparable in size to the thickness of the film. This causes a huge stress relaxation to occur. The excess volume associated with the large number of grain boundaries in the as-deposited state is removed during heating and this densification relaxes the compressive stress in the film. This mechanism of stress relaxation does not involve plastic deformation of the film; it is the result of a structural transformation within the film. The stress relaxation associated with grain growth occurs only during the first heating cycle. For subsequent heating and cooling cycles, the grain structure remains essentially constant and the stress-temperature curves are dominated by elastic and plastic deformation of the film. The stress changes associated with grain growth could cause the stress to change sign. If the tensile stress in the as-deposited state were greater, then grain growth would have caused the stress to go back into tension in the temperature range of 250-300°C.

In some cases the stress-temperature relations are almost completely dominated by structural transformations that occur upon heating and cooling. An example is shown in Fig. 23 for a chemically vapor deposited (CVD) film of WSi_2 on Si [26]. This film is essentially amorphous in the as-deposited state and is in a state of biaxial tension. The stress decreases slightly on heating due to the differences in thermal expansion of the film and substrate. At about 500°C the film crystallizes and densifies, causing the tensile stress in the film to increase abruptly. At still higher temperatures the film begins to deform plastically and this causes the stress to relax. Heating to 900°C leads to a very stable microstructure that does not change during subsequent heating and cooling cycles. After the structure has been stabilized in this way, the film deforms in a purely reversible, elastic manner on subsequent heating and cooling.
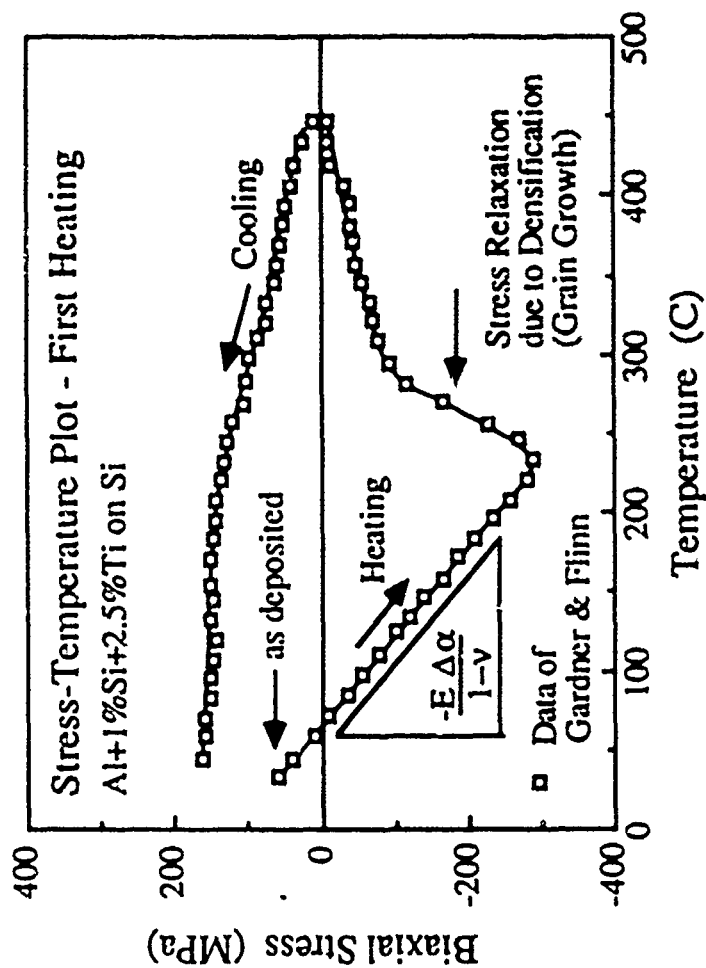
28

Figure 22. Stress - temperature plot for an Al+1%Si+2.5%Ti film on a Si substrate [27]. The stress observed during the first heating and cooling cycle are shown. The large stress relaxation observed near 250°C is caused by densification of the film via grain growth.
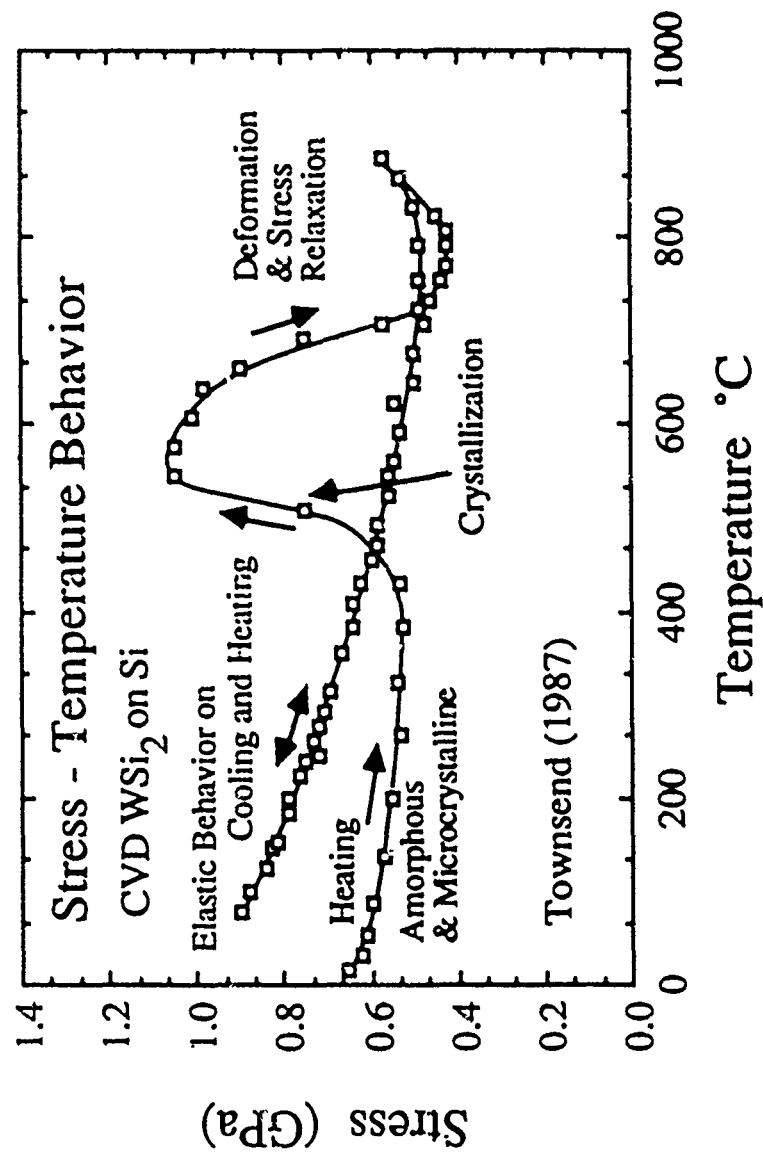
Figure 23. Stress - temperature behavior for an initially amorphous film of CVD WSi2 on a Si substrate. The large peak is caused by crystallization and deformation of the film. A purely elastic response is observed for subsequent heating and cooling cycles.

Strengths of Thin Films on Substrates

We now consider the mechanisms of plastic deformation in thin films on substrates. We wish to understand the factors that control the strengths of these thin film materials. We note that materials in thin film form exhibit much higher strengths than their bulk counterparts. For example, the data in Fig. 20 indicate that the room temperature yield strength of pure aluminum is about 280 MPa in thin film form. The compressive flow stress at high temperatures is about 100 MPa. These strengths are much higher than those found for bulk pure aluminum.

The high strengths of thin films on substrates can be attributed partly to the fine microstructures found in these materials. As noted above, even after annealing, the grain size typicall·· does not exceed the thickness of the film. As a consequence, significant grain size strengthening occurs naturally in these thin film materials. In addition, the dislocation densities in thin films can be quite high and this also contributes to the high strengths of these materials. However, these microstructural factors cannot fully account for the high strengths of thin films on substrates. Much of the strength can be attributed to the fact that the substrate constrains dislocation motion in the film. Any oxides that may be present on the surface of the film can also constrain the movements of dislocations within the film. The effects of these constraints on dislocation motion are shown in Fig. 24. Here we envision a single crystal film attached to a substrate; the film is assumed to have a non-deformable oxide on its surface. A single dislocation moving within the film must leave a dislocation line at each interface as it moves. In this respect the dislocation motion is analogous to the movement of threading dislocations in epitaxial films. We can use the approach developed for threading dislocations to describe the motion of dislocations in thin films and this will permit us to understand the high strengths of thin films on substrates.

The Misfit Dislocation Model

Following the approach developed by Freund [17], we consider the work done by the biaxial stress in the film when the dislocation travels a unit distance. For the geometry shown in Fig. 24 this may be expressed as

$$W_{layer} = \frac{\cos \varphi \cos \lambda}{\sin \varphi} \sigma bh \quad , \tag{34}$$

Figure 24.   Dislocation motion in an oxidized film on a non-deformable substrate. Misfit dislocations are produced at the two interfaces.

where σ is the biaxial stress in the film, b is the Burgers vector and h is the thickness of the film. Here we have retained the notation of eqn. (19) and used the subscript "layer" to indicate the work done by the stress in the film. In order for the dislocation to move, the work done by the stress in the film must be equal to or greater than the energies of the two dislocations left at the film-substrate and film-oxide interfaces, which we express as

$$W_{dislocations} = \frac{b^2}{4\pi(1-\nu)} \left[\frac{2\mu_f \mu_s}{(\mu_f + \mu_s)} \ln\left(\frac{\beta_s h}{b}\right) + \frac{2\mu_f \mu_o}{(\mu_f + \mu_o)} \ln\left(\frac{\beta_o t}{b}\right) \right] \quad , \qquad (35)$$

where μ_f, μ_s and μ_o are the elastic shear moduli of the film, substrate and oxide, respectively, h and t are the thicknesses of the film and oxide, and $\beta_s = 2.6$ and $\beta_o = 17.5$ are numerical constants. By equating W_{layer} to $W_{dislocations}$ we obtain an expression for the minimum biaxial stress needed to move the dislocation in the film

$$\sigma = \frac{\sin \varphi}{\cos \varphi \cos \lambda} \frac{b}{2\pi(1-\nu)h} \left[\frac{\mu_f \mu_s}{(\mu_f + \mu_s)} \ln\left(\frac{\beta_s h}{b}\right) + \frac{\mu_f \mu_o}{(\mu_f + \mu_o)} \ln\left(\frac{\beta_o t}{b}\right) \right] \quad . \qquad (36)$$

This is an expression for the minimum biaxial yield strength of a single crystal thin film on a substrate. No obstacles to dislocation motion, such as other dislocations, point defects or grain boundaries, have been assumed and no friction stress for dislocation motion has been considered. If an oxide is not present on the surface of the film, the second term in the brackets should be deleted. The most important feature of this result is the prediction that the yield strength of the film depends inversely on the thickness of the film. This helps to explain why very thin films have high yield strengths.

Comparison of the Model with Experiment

The predictions of eqn. (36) can be compared with the measured strengths of thin films on substrates. Kuan and Murakami [28] have used x-ray techniques to measure the biaxial yield strengths of polycrystalline films of Pb on silicon at 4.2 K as a function of film thickness. Some of their results are shown in Fig. 25. The films exhibit a strong <111> texture and have a large grain size. The predictions of eqn. (36) are shown as a solid line in the figure. The following constants were used in this calculation: $\mu_f = 13.6$ GPa, $\mu_s = 66.5$ GPa, $\mu_o = 64$ GPa, b=0.35 nm, $\nu = 0.376$, t=5 nm and $\sin\varphi/\cos\varphi\cos\lambda = 3.464$. The model gives a good account of strengths of these films. Doerner, Gardner and Nix [29] have also measured the biaxial strengths of thin films
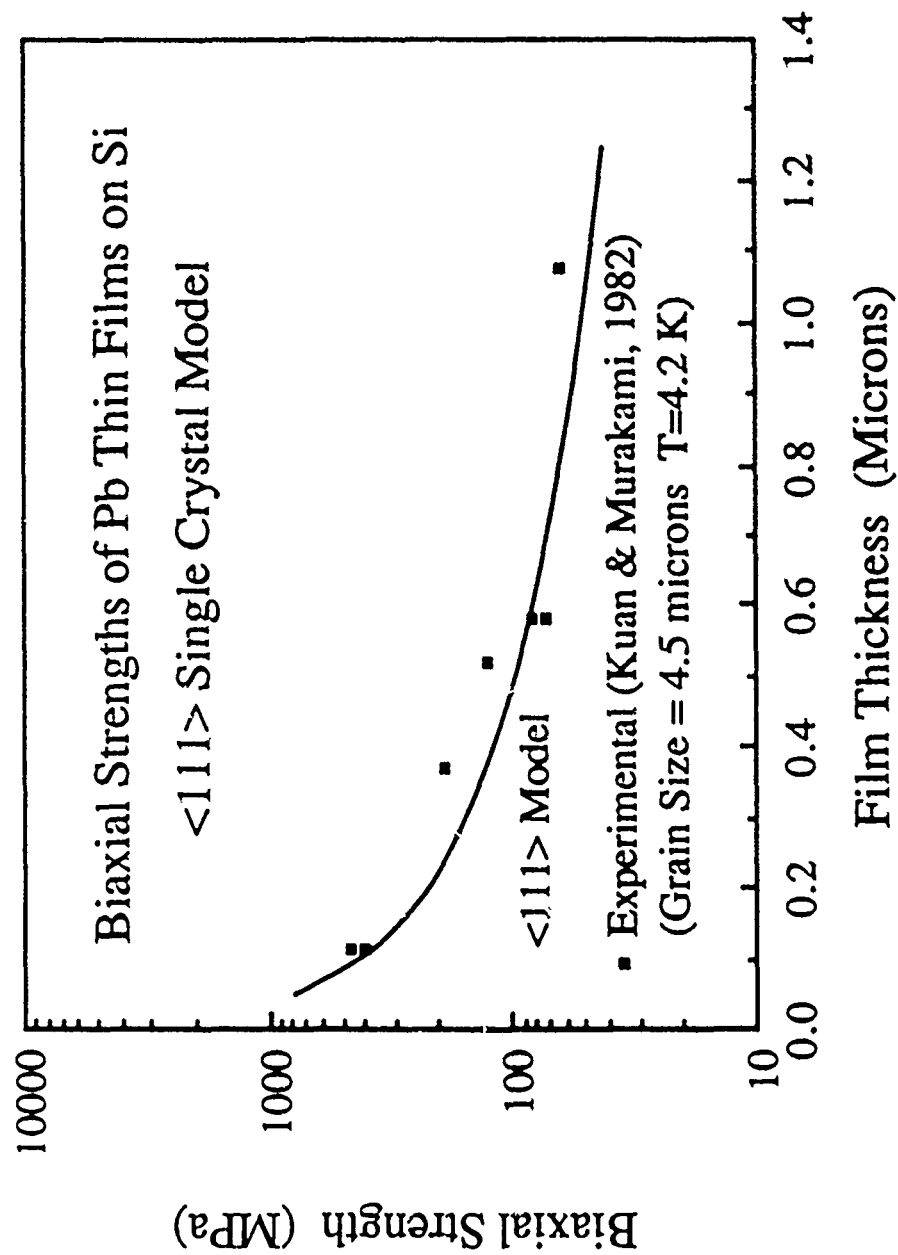
30

Figure 25.   Biaxial strengths of Pb thin films on Si substrates as a function of film thickness [28].   Predictions of the misfit dislocation model for film strength are shown for comparison.

of pure aluminum as a function film thickness using the substrate curvature technique described earlier. The stress-temperature curves for two different film thicknesses are shown in Fig. 26. We note that the strength of the thinner film is greater at both high and low temperature. This suggests that an athermal strengthening mechanism is primarily responsible for the strength of the film. The mechanism described earlier is essentially athermal as indicated by the absence of a strongly temperature dependent term in eqn. (36). The predictions of eqn. (36) are compared with the experimental results of Doerner et al. [29] in Fig. 27. The curve labeled σ_0 represents the prediction of eqn. (36) using the following constants: $\mu_f=24.8$ GPa, $\mu_s=66.5$ GPa, $\mu_0=178.9$ GPa, $b=0.286$ nm, $v=0.31$ and $\sin\varphi/\cos\varphi\cos\lambda=3.464$. It is evident that the measured yield strengths are greater than the predictions of eqn. (36), especially at the larger film thicknesses. The discrepancy is due to the omission of grain size strengthening in the above model. Doerner [30] measured the grain sizes of the aluminum films with transmission electron microscopy and found a systematic variation of grain size with film thickness. She found the grain size to be about 1.3h, where h is the thickness of the film. Using this relation together with the Hall-Petch coefficient of $k_y=6\times10^4$ N/m$^{3/2}$ reported for bulk aluminum by Hansen [31] and Armstrong [32], we can find the grain size contribution to the strength. This contribution is shown as a solid line labeled $k_y d^{-1/2}$ in Fig. 27. By adding this contribution to σ_0 we obtain a prediction that is in reasonable agreement with the data. A grain size contribution is negligable for the Pb data of Fig. 25 because the grain size is so much larger than the film thickness.

We have argued that the constraint of the substrate makes a significant contribution to the strength of a thin film on a substrate. One way to check this prediction is to compare the strengths of thin films on substrates with the strengths of free-standing thin films. A comparison of this kind is shown is Fig. 28. Here the yield strength data of Doerner et al.[29] for pure aluminum films on silicon are compared with biaxial yield strengths of free standing thin films of Al+1%Si reported by Griffen at al. [33-34]. Griffen's data were obtained by bulge testing of free standing films, so they do not include the effect of the constraint of the substrate. All of the data are plotted according to the Hall-Petch relation. The data of Griffen et al.[33-34] were obtained for films with a constant thickness of 1 μm, whereas the data of Doerner et al.[29] were obtained for films of various thicknesses. Thus, the Hall-Petch slope for the Doerner data is somewhat misleading because much of the strengthening is actually caused by the constraint of the substrate. The Hall-Petch slope for bulk polycrystalline aluminum is shown for comparison. We note that the strengths of the aluminum films bonded to the silicon substrate are much greater than the strengths of the free standing films. The constraining effect of the substrate makes these films even stronger than bulk aluminum of comparable grain size. These comparisons support the idea that much of the strength of thin films on substrates is caused by the constraint of the substrate.
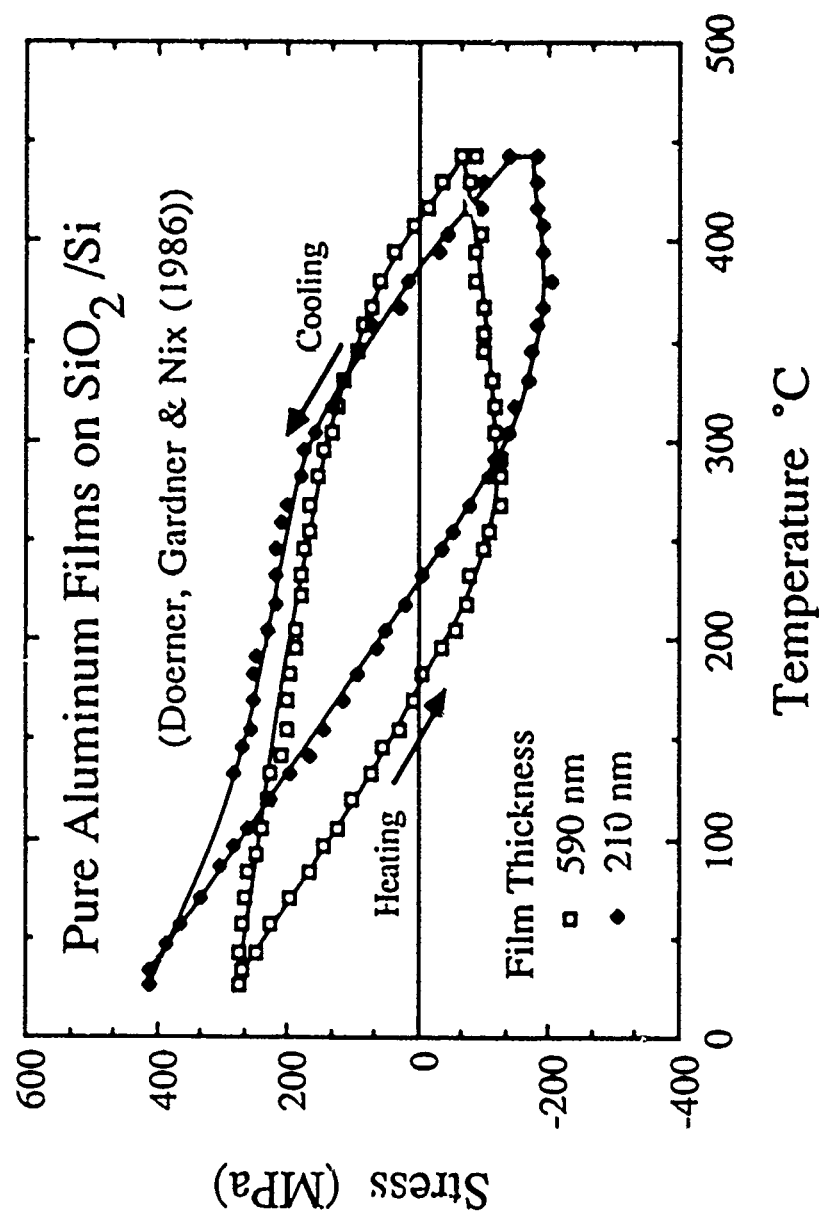
31

Figure 26. Stress - temperature plots for thin films of Al on a Si substrate [29]. The results show that the thinner film is stronger than the thicker film at both low and high temperatures.
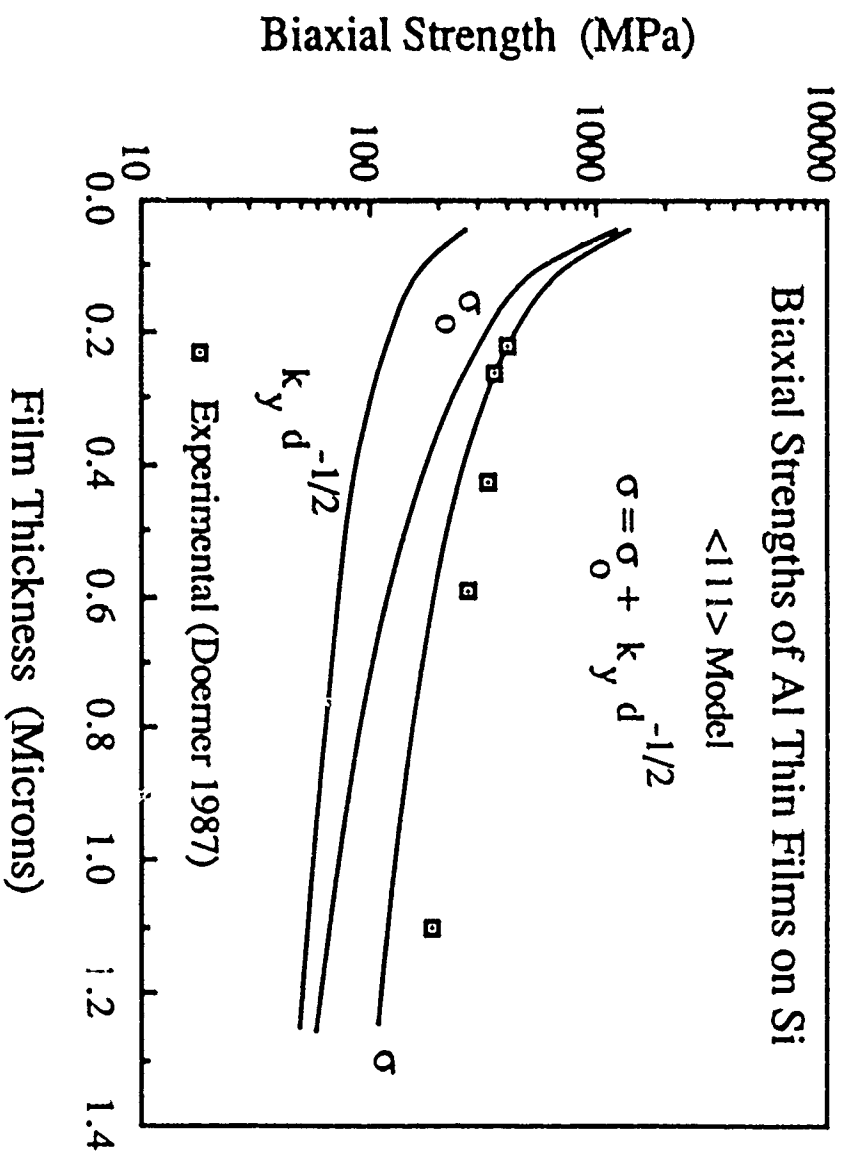
Figure 27. Biaxial strengths of Al thin films on Si substrates [29]. The predictions of the misfit dislocation model for film strength are shown for comparison. Hall-Petch strengthening must be included to account for the high strengths of the films.
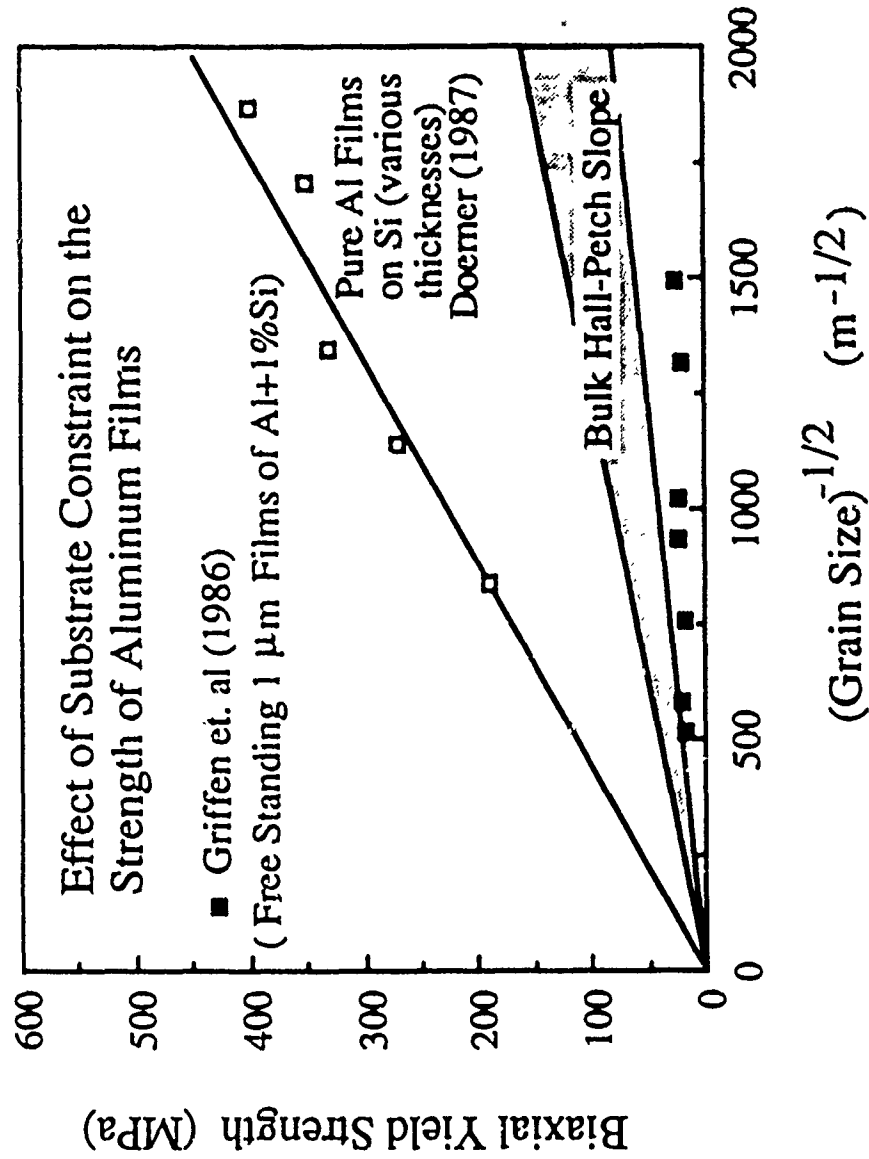
Figure 28. Comparison of the strengths of the Al films on Si substrates [30] with the biaxial strengths of free standing films of Al+1% Si [33-34], in the form of a Hall-Petch plot. The constraint of the substrate causes the films bonded to the substrate to be much stronger than the free standing films.

VI. Sub-Micron Indentation Testing Techniques

In conventional studies of the mechanical behavior of materials it is common practice to create test samples of the material in question, subject them to forces or displacements and measure the corresponding response. Usually simple stress states such as tension, compression and torsion are used to permit the calculations of stress and strain from the measured forces and displacements. Such an approach cannot be used to study the mechanical properties of thin films on substrates, because the thin films of interest are typically 1µm in thickness or less, and are bonded to substrates. Sub-micron indentation represents an alternative approach. By measuring the forces needed to create small indentations in thin films on substrates it is possible the obtain information about the elastic and plastic properties of the film material. An inherent complication of this approach is that the stresses and strains produced by indentation are non-uniform within the test sample. Therefore only nominal values of the stresses, strains and strain rates can be determined in such tests. Nevertheless, sub-micron indentation gives some information about the mechanical properties of thin films that cannot be obtained easily in any other way.

Depth-Sensing Indentation Instruments

Although indentation hardness testing has been in widespread use for more than a century, high resolution (sub-micron) indentation instruments have become available only recently. Most of the instruments that are available have been built by individual investigators to meet specific research requirements [35-43]. In the traditional approach to the determination of hardness, a fixed load is imposed on a diamond indenter and the resulting indentation is observed. The hardness is taken to be the load divided by the projected area of the permanent indentation. This approach has been adopted by Bangert et al. [40] in their development of a microindentation attachment used in conjunction with a scanning electron microscope. Other instruments that do not require the direct observation of the indentation have also been developed [35-39,41-43]. These instruments are called depth-sensing because both the load on the indenter and the displacements associated with indentation are measured in the course of indentation. An instrument of this kind, called the Nanoindenter, was developed by Pethica and Oliver [37,43] and is commercially available. Stanford University has such a devive and results obtained with that machine will be presented here. The basic characteristics of this instrument and the mechanical properties that can be measured with it are described below. Other instruments of this kind have been built by other investigators. Most of these are, like the Nanoindenter, load-controlled instruments in which the force on the indenter is imposed and the resulting displacement of the indenter into the material is measured [35-39]. These instruments can be regarded as "soft" testing machines in the sense that

32

the load on the indenter is not changed by the penetration of the indenter into the material, but rather is controlled by the loading mechanism. Such instruments are analogous to creep machines in ordinary mechanical testing. Another form of microindentation can be done under displacement control. Li and Wu and their collaborators [41-42] have developed instruments that allow the displacement of the indenter into the material to be imposed and the resulting load to be measured. Such instruments are analogous to "hard" testing machines that are typically used in tensile testing. They are also well suited for the study of stress relaxation in indentation [41,44].

The microindentation experiments described here are characterized by the condition that the indenter makes increasing contact with the material during indentation. A different kind of mechanical response is observed when the indenter is flat-ended and has the shape of a right circular cylinder. For such an indenter the projected contact area does not change as the indenter is pushed into the material. This response has been called impression testing by Li, Hu and Rath [45-47] and has been used to study the mechanical properties of small volumes of material. However, the impression tools used in such instruments are typically too big to be used in the study of most thin films on substrates.

The Nanoindenter

A schematic diagram of the Nanoindenter is shown in Fig. 29 [37,48]. A diamond indenter is fixed to the end of a loading shaft that is suspended on delicate leaf springs. The leaf springs are compliant in the loading direction but stiff in the transverse direction. When the indenter is not in contact with the material, the shaft is supported both by the suspending springs and a force that is delivered by the coil and magnet assembly at the top of the shaft. Movement of the indenter toward the sample is accomplished by reducing the force supplied by the magnet-coil assembly. The smallest load increment that can be made with the Stanford instrument is about 0.25 μN and the maximum force that can be exerted is 120 mN. Motion of the indenter toward the sample is measured by a capacitance displacement gage with a displacement resolution of 0.4 nm. With this resolution it is possible to make indentation measurements in materials with total indentation depths as shallow as about 20 nm. As noted above, the indentation made by the indenter is not observed, but rather is determined from the measured depth of indentation and the known shape of the indenter. The sample is mounted on an x-y table that allows indentations to be precisely located with respect to each other on the same surface. The position of the first indentation of an array can be placed within a few micrometers of a given location; the remaining indentations of the array can be placed with respect to the first one with a resolution of about 0.5 μm..

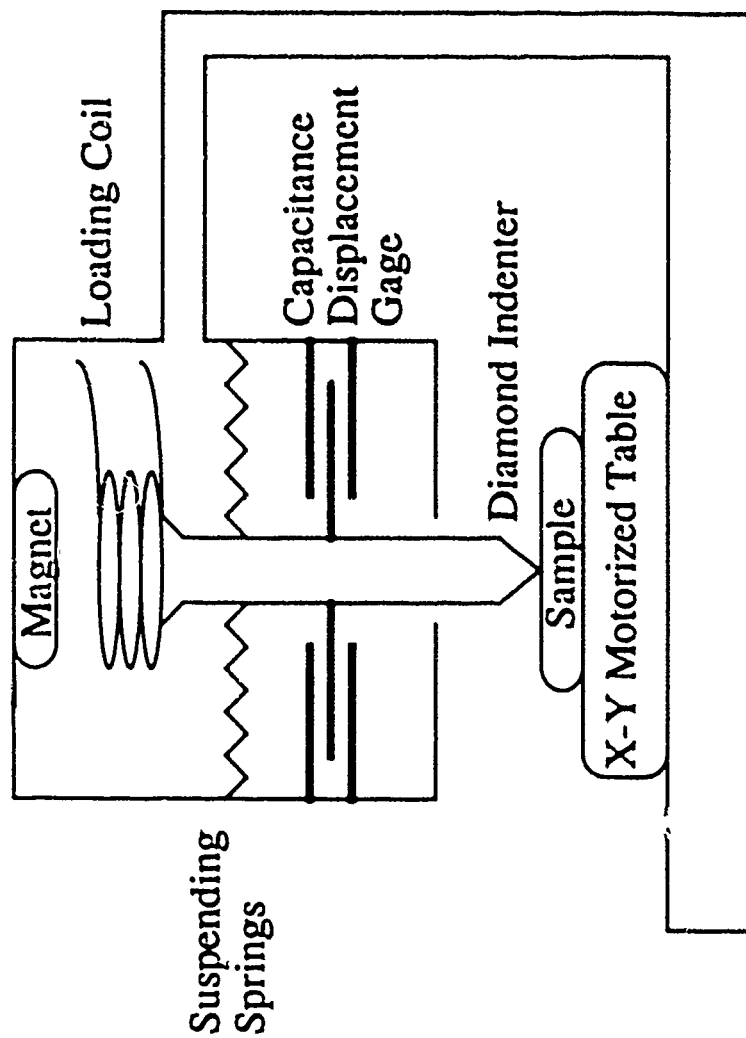33

THE NANOINDENTER

Figure 29.   Schematic diagram of the Nanoindenter.

Indenters of any shape can be used in microindentation but since small contact areas are needed, the best shape is the three faced Berkovich indenter shown in Fig. 30. Because any three non-parallel planes intersect at a single point, it is relatively easy to grind a sharp tip on an indenter if the three-faced Berkovich geometry is used. The angles of the faces are fixed so that the nominal relationship between the area and depth of indentation is the same as for the Vickers indenter. Invariably, the tip of the indenter is rounded so that the ideal geometry is not maintained near the tip. As shown schematically in Fig. 31, the actual depth of indentation produces a larger contact area than would be expected for an indenter with an ideal shape. We use the effective depth h_{eff} to indicate the depth of indentation that would be needed for an ideally shaped indenter to produce the same contact area as the non-ideal tip for a plastic depth h_p. For the Berkovich (and Vickers) indenter, the nominal shape is characterized by

$$\sqrt{A} = kh_{eff} = 24.5h_{eff} \quad , \qquad (37)$$

where A is the contact area between the indenter and the material. For the diamond indenter used in most of the Stanford tests, the radius of curvature at the tip is about 300 nm.

Indentation Mechanical Properties

In a typical experiment, the tip of the indenter is moved toward the surface of the sample by gradually increasing the load on the indenter shaft. With a constant loading rate of $1 \mu N/s$, the tip of the indenter travels downward at a velocity of about 20 nm/s. When the tip contacts the surface, its velocity drops below 1 nm/s and the computer records the displacement of the tip. This is the point at which the indentation experiment begins. The loading rate is subsequently adjusted to maintain the descent velocity between 3 and 6 nm/s. A typical indentation loading curve is . ~ n schematically in Fig. 32 where the load on the indenter is plotted against the depth of indentation. The process of indentation causes both elastic and plastic deformation to occur, as shown in Fig. 33. The tip of the indenter is usually sharp enough to cause plastic flow to occur from the very beginning of the test, so that an elastic-plastic transition typical of an ordinary tensile test is not observed. The slope of the loading curve increases with depth of indentation because the indenter makes increasing contact with the material during indentation. The hardness of the material can be found by dividing the indentation force by the contact area at each point along the loading curve. This permits a measurement of the hardness as a function of the indentation depth. If the load is held constant, the indenter will continue to sink into the material in a time-dependent manner as creep flow under the indenter occurs. This allows a study of the creep properties of very small
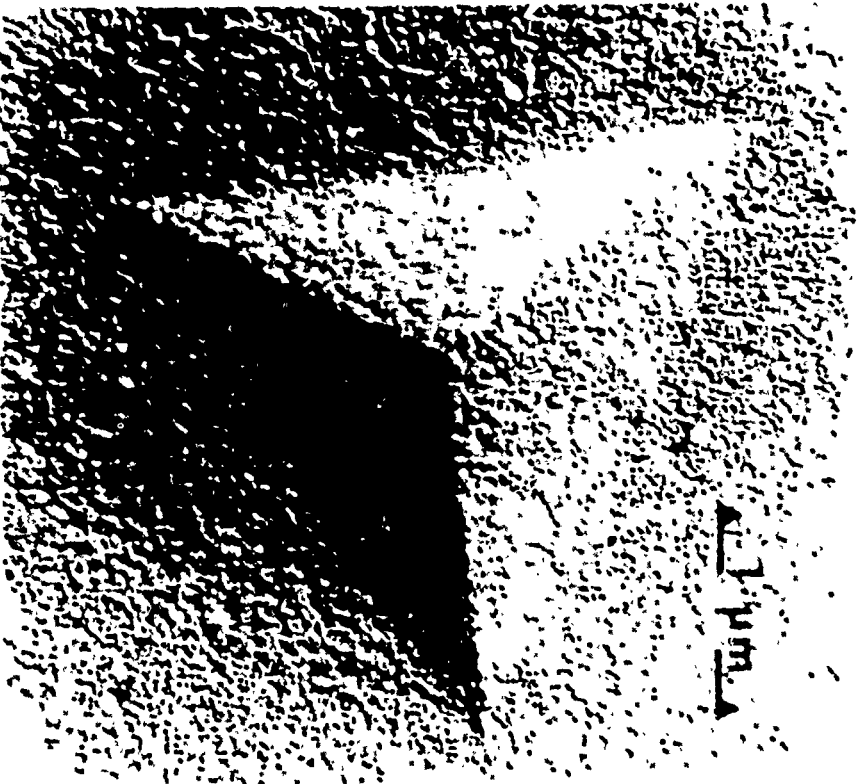
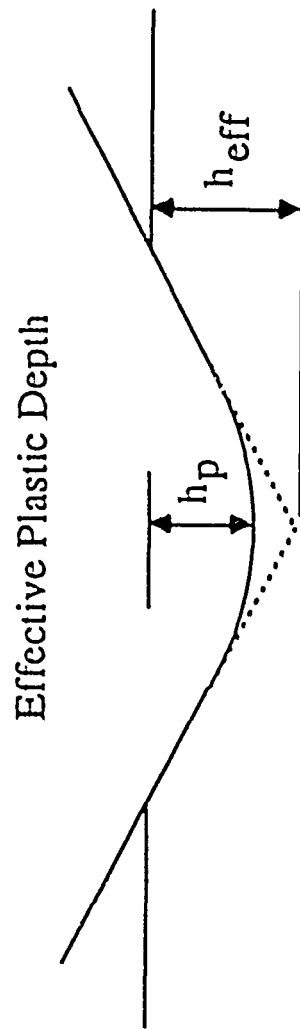Figure 30. Replica transmission electron micrograph of the Berkovich indentation.

Figure 31. Definitions of the plastic depth and the effective depth for an indenter with a non-ideal shape. The effective depth is the depth at which an ideal indenter would create the same contact area as the real indenter.
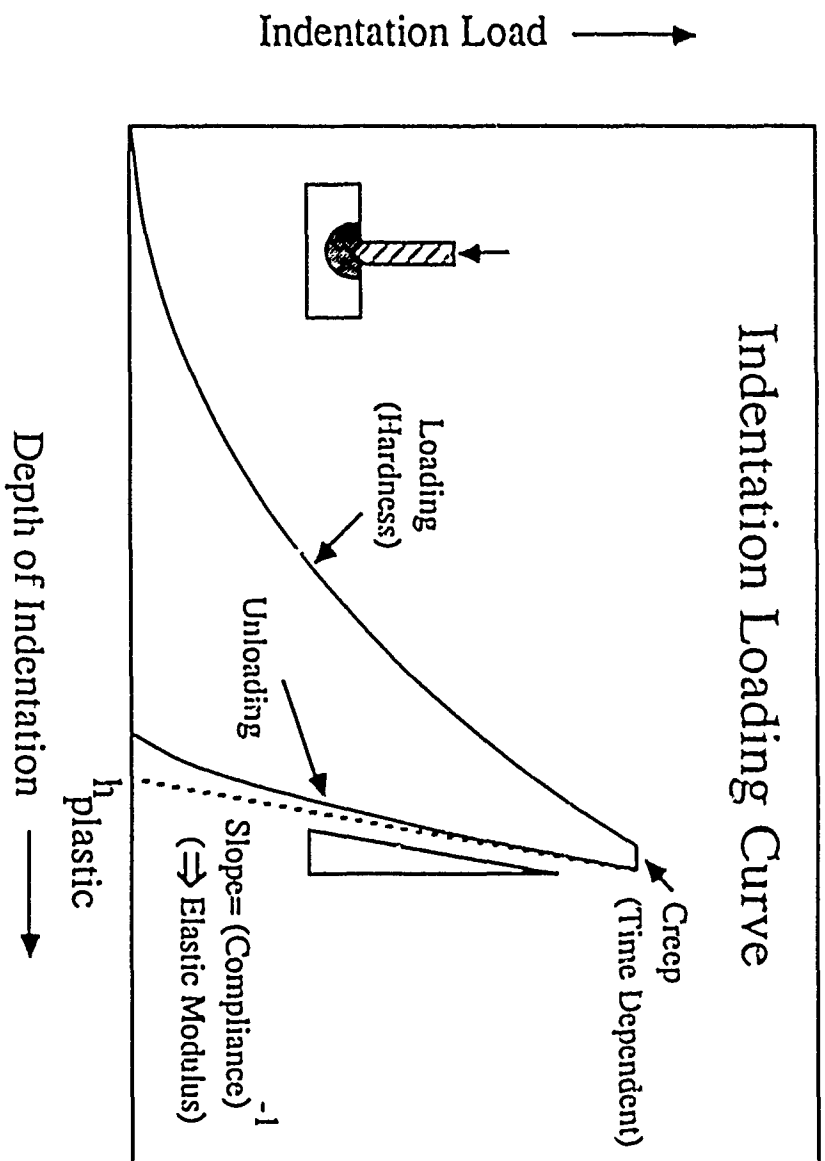
Figure 32. Indentation loading and unloading curve. Hardness, creep and elastic properties can be obtained from these data.
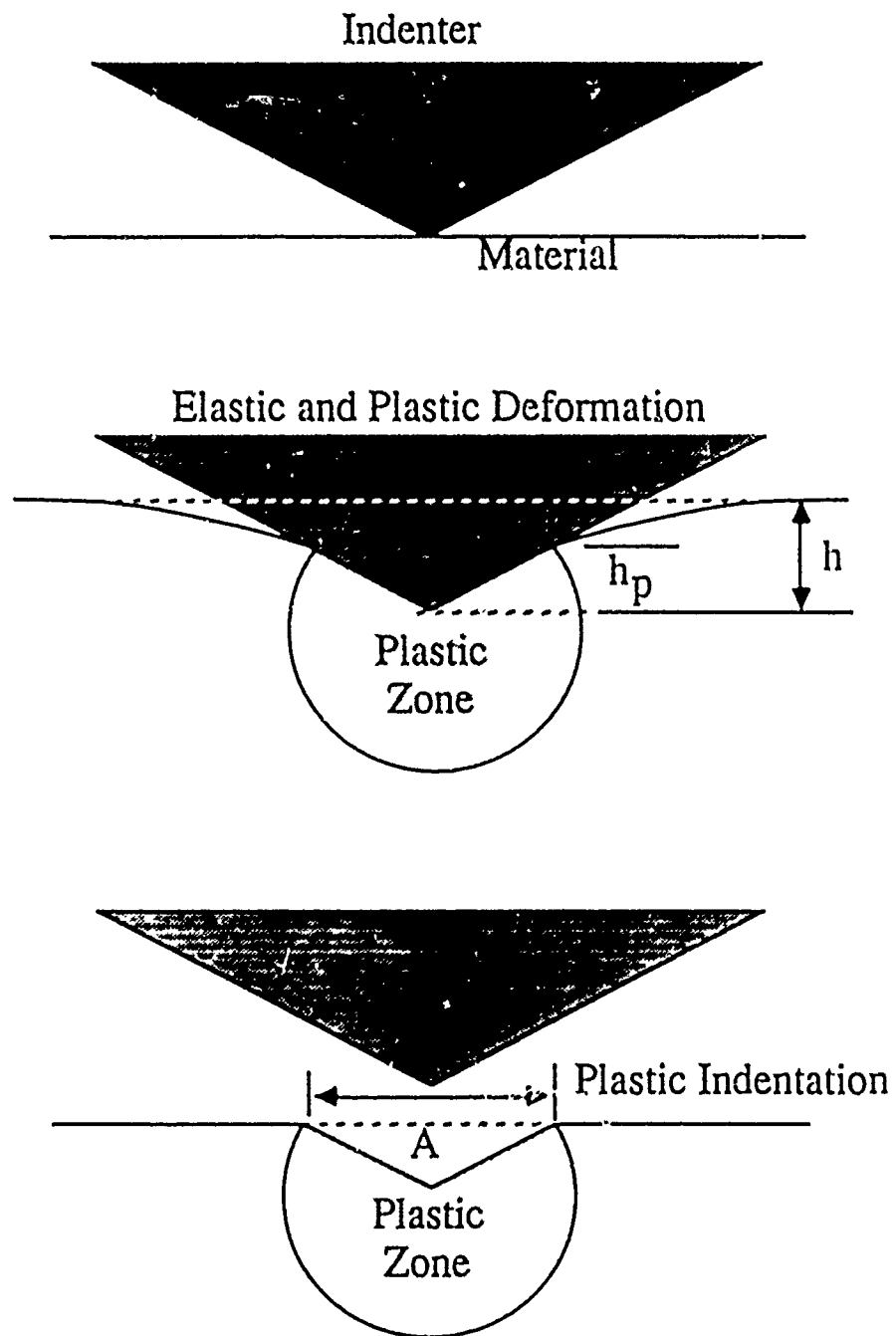
The figure shows an "Indentation Loading Curve" plotting Indentation Load versus Depth of Indentation, with labels:
- Loading (Hardness)
- Unloading
- Creep (Time Dependent)
- Slope = (Compliance)$^{-1}$ ($\rightarrow$ Elastic Modulus)
- $h_{plastic}$

Figure 33. Illustration of elastic and plastic deformation associated with indentation.

volumes of material. During unloading, the indenter is pushed back out of the material by the elastic restoring forces in the system. This effect can be used to determine elastic properties. We begin our discussion of mechanical properties by showing that the elastic properties of materials can be determined by analyzing the unloading portion of the indentation curve.

Elastic Properties

From the slope of the unloading curve it is possible to determine the elastic modulus of the material being indented, provided the contact area and the elastic properties of the indenter are known. Loubet et al. [38] showed that the slope of the unloading curve could be modeled by treating the indenter as a flat-ended elastic punch causing displacements to occur in an elastic half space. According to such a model the unloading slope is given by

$$\frac{dP}{dh} = \beta E^* \sqrt{A} \qquad , \tag{38}$$

where P is the load, h is the depth of indentation, A is the contact area, β is a constant that depends slightly on the shape of the indenter and E^* is an effective modulus for the system defined by

$$\frac{1}{E^*} = \frac{1 - v_o^2}{E_o} + \frac{1 - v^2}{E} \qquad , \tag{39}$$

where E_o and v_o are Young's modulus and Poisson's ratio of the indenter and E and v are the same elastic properties of the material being indented. For a cylindrical punch, Sneddon [49] showed that the constant β is $2/\pi^{1/2}$ King [50] later showed that the constant is only slightly different for other shapes:

Circle $\qquad \frac{\sqrt{\pi}}{2}\beta = 1.0$

Square $\qquad \frac{\sqrt{\pi}}{2}\beta = 1.011$

Triangle $\qquad \frac{\sqrt{\pi}}{2}\beta = 1.034$

The non-linearity of the unloading curve indicates that the contact area decreases during the course of unloading. In order to determine the contact area at the point of maximum load, the slope of the unloading curve is extrapolated to zero load and the abscissa is read as the plastic depth, as

shown in Fig. 32. The plastic depth measured in this way, together with a knowledge of the shape of the indenter obtained from calibration experiments permits a determination of the contact area. This contact area is related to the effective depth h_{eff} using eqn. (37). Because the compliance of the Nanoindenter itself contributes to measured displacements at high loads, it is preferable to write eqn. (38) in terms of the compliance:

$$\frac{dh}{dP} \approx C_m + C_{indent}$$

$$\frac{dh}{dP} = C_m + \frac{1}{\beta\sqrt{A}}\cdot\frac{1}{E^*} \quad , \tag{40}$$

where C_m is the compliance of the machine and C_{indent} is the compliance of the indenter and material. Using eqn. (37), this expression can be written in terms of the effective depth h_{eff} as:

$$\frac{dh}{dP} = C_m + \frac{1}{\beta k}\frac{1}{E^*}\frac{1}{h_{eff}} \quad . \tag{41}$$

We see that the measured compliance should vary linearly with the reciprocal of the effective indentation depth. As shown schematically in Fig. 34, the slope of a plot of compliance versus the reciprocal of the effective depth gives a measure of the effective modulus of the system. The elastic properties of materials can be measured in this way. Some results for various materials are given in Fig. 35 and compared with elastic constants taken from the literature. The agreement is very good for soft materials such as aluminum. For harder materials the indentation experiments give moduli that are higher than expected. We believe this is caused partly by the high pressure under the indenter. Some densification of SiO_2 may occur under the indenter and this may cause an increase in the elastic modulus. The pressure can also increase the elastic modulus of the material just under the indenter through non-linear elastic effects. This may contribute to higher values of elastic modulus found for silicon and tungsten.

It should be noted that the procedure outlined here for the measurement of elastic properties can be applied to hard materials that show little or no plastic indentation. In the limit of no plastic deformation (with no friction), the unloading curve would coincide exactly with the loading curve and all of the displacements associated with indentation would be elastic. Under these limiting conditions the slope of the curve at any point would indicate the contact area at that point. Pethica and Oliver [43] showed that purely elastic deformation is observed for small indentations in electropolished single crystals of tungsten. They assumed that for small indentations the contact

36

Schematic Compliance Plot

Elastic Compliance, $\dfrac{dh}{dP}$

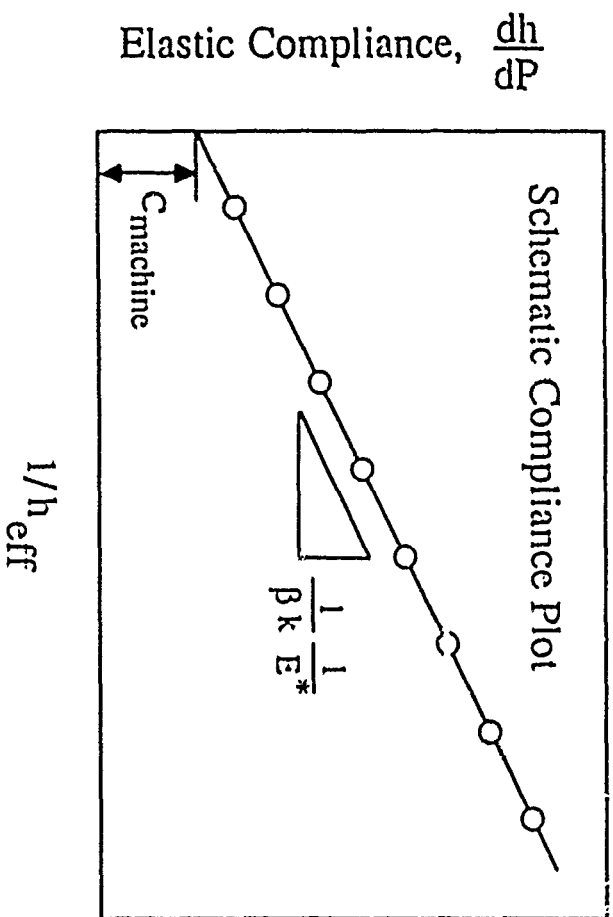$$\frac{1}{\beta k}\frac{1}{E^*}$$

$C_{machine}$

$1/h_{eff}$

Figure 34. Schematic plot of the elastic compliance as a function of the reciprocal of the effective depth for an indentation system. The intercept gives the elastic compliance of the machine and the slope permits a calculation of the elastic properties of the material being indented.
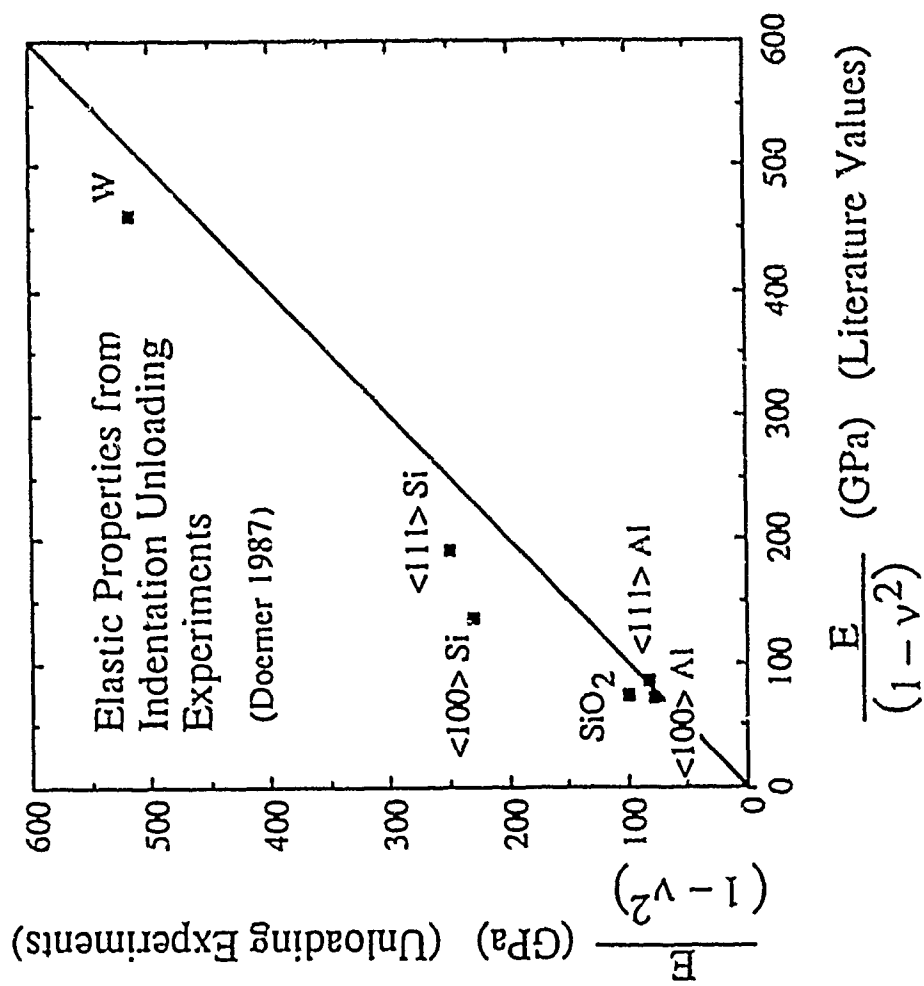
Figure 35. Elastic properties of several materials obtained by indentation unloading experiments [48]. The results are compared with literature values of the elastic properties. The agreement is best for soft metals such as Al.

pressures are insufficient to cause dislocation nucleation and motion to occur in tungsten. They found the elastic properties to be well described by eqn. (38) in this regime. They also showed that for larger indentations, plastic flow occurs as a sudden increase in indenter displacement at a critical contact pressure. This event is assumed to coincide with the nucleation of dislocations under the indenter. That a critical contact pressure is needed for dislocation nucleation had been confirmed in earlier work on sapphire by Page, Oliver and McHargue [51].

The elastic properties of thin films on substrates can also be measured by indentation [48,50,5?]. It is evident that the indentation compliance should depend on the depth of indentation compared to the thickness of the film. It is expected that for very large indentations the compliance should be dominated by the elastic properties of the substrate, whereas for small indentations the compliance should be dominated by the elastic properties of the film. Doerner and Nix [48] proposed a simple phenomenological relation to describe the relative contributions of the film and substrate to the indentation compliance. Their expression for the indentation compliance (after subtracting the machine compliance from the measured compliance) is

$$\left(\frac{dh}{dP}\right)_t = \frac{1}{\beta k}\left(\frac{t}{h_{eff}}\right)\left\{\frac{1-v_o^2}{E_o} + \frac{1-v_s^2}{E_s}\exp\left(-\frac{\alpha}{k}\frac{t}{h_{eff}}\right) + \frac{1-v_f^2}{E_f}\left[1-\exp\left(-\frac{\alpha}{k}\frac{t}{h_{eff}}\right)\right]\right\} \quad (42)$$

where E_s and v_s stand for the elastic properties of the substrate, E_f and v_f stand for the elastic properties of the film, t is the film thickness and α is a parameter that characterizes the transition from film to substrate control. The expression shows that the product of the compliance and the film thickness depends uniquely on the ratio of the film thickness to the effective depth. Doerner and Nix [48] showed that this relation could be used to describe the elastic properties of thin films of tungsten on silicon if α was chosen to be 1.24. Their results are shown in Fig. 36. The solid lines indicate the compliances of bulk silicon and tungsten. For small values of t/h_{eff} (i.e. large indentation depths) the compliance data coincides with the compliance of the silicon substrate, as expected. For large values of t/h_{eff} (i.e. small indentation depths) the compliance data approaches the compliance of bulk tungsten. This the experiment shows how it might be possible to determine the elastic properties of thin films on substrates from the composite modulus of the film and substrate.

Recent theoretical work by King [50] has confirmed that eqn. (42) gives a good account of the indentation elastic properties of thin films on substrates. However, King's work indicates that the parameter α is not constant but varies slightly with the relative depth of indentation.
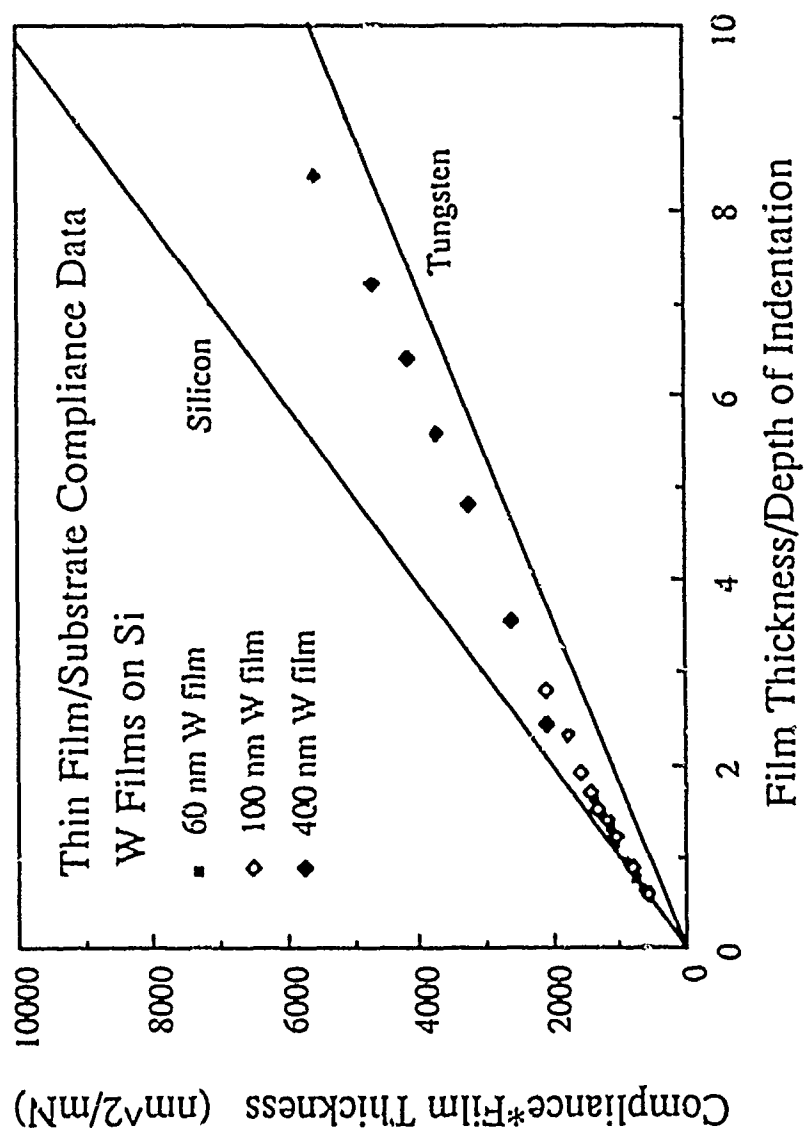
37

Figure 36. Elastic properties of thin films of W on Si substrates as determined from the measurement of unloading compliances [48]. The solid lines which indicate the measured compliances of bulk W and Si also represent the limiting compliances for the thin film - substrate measurements.

Hardness

As noted earlier, hardness is usually defined as the load divided by the projected area of the indentation left by the indenter. The problem of measuring the hardness of thin films on substrates is to avoid the influence of the substrate in the measurement. For the case of metal films on silicon, the substrate is typically much harder than the film. In this case the indentation used to make the hardness measurement must be small compared to the film thickness. The displacement resolution of the Nanoindenter permits measurements of this kind to be made.

Hardness measurements for a pure aluminum film, 1 μm thick, deposited onto an oxidized silicon substrate are shown in Fig. 37 as a function of the depth of indentation. The smallest indentations give hardnesses that coincide with the hardness of single crystal aluminum. This is to be expected as the smallest indentations probably occur within a single grain of aluminum. The measured hardness increases slightly with the depth of indentation, as expected. The harder substrate causes the measured hardness to increase when the plastic zone under the indenter reaches the film-substrate interface. An elastic-plastic finite element analysis of this indentation problem has been made by Bhattacharya and Nix [52]. One of their results is shown in Fig. 37. This continuum analysis shows that the hardness should increase gradually with increasing depth of indentation, especially for indentation depths greater than about 20% of the thickness of the film. The calculated hardness does not reach the substrate hardness even for large indentation depths because most of the indenter still rests on the soft aluminum film.

The good agreement between the finite element analysis and the measured hardness does not extend to thinner films. As shown in Fig. 38, for thinner films the measured hardness increases much more abruptly with indentation depth than the finite element analysis predicts. Several factors are believed to be responsible for this effect. The thinner films are known to have much finer grain sizes, so that a Hall-Petch hardening effect is to be expected. In addition, as discussed earlier in connection with misfit dislocations, dislocation motion in thin films requires higher shear stresses than in the bulk because of the constraining effect of the non-deformable substrate. This too may help to explain why the hardness of very thin films is so much greater than the predictions of the continuum analysis.
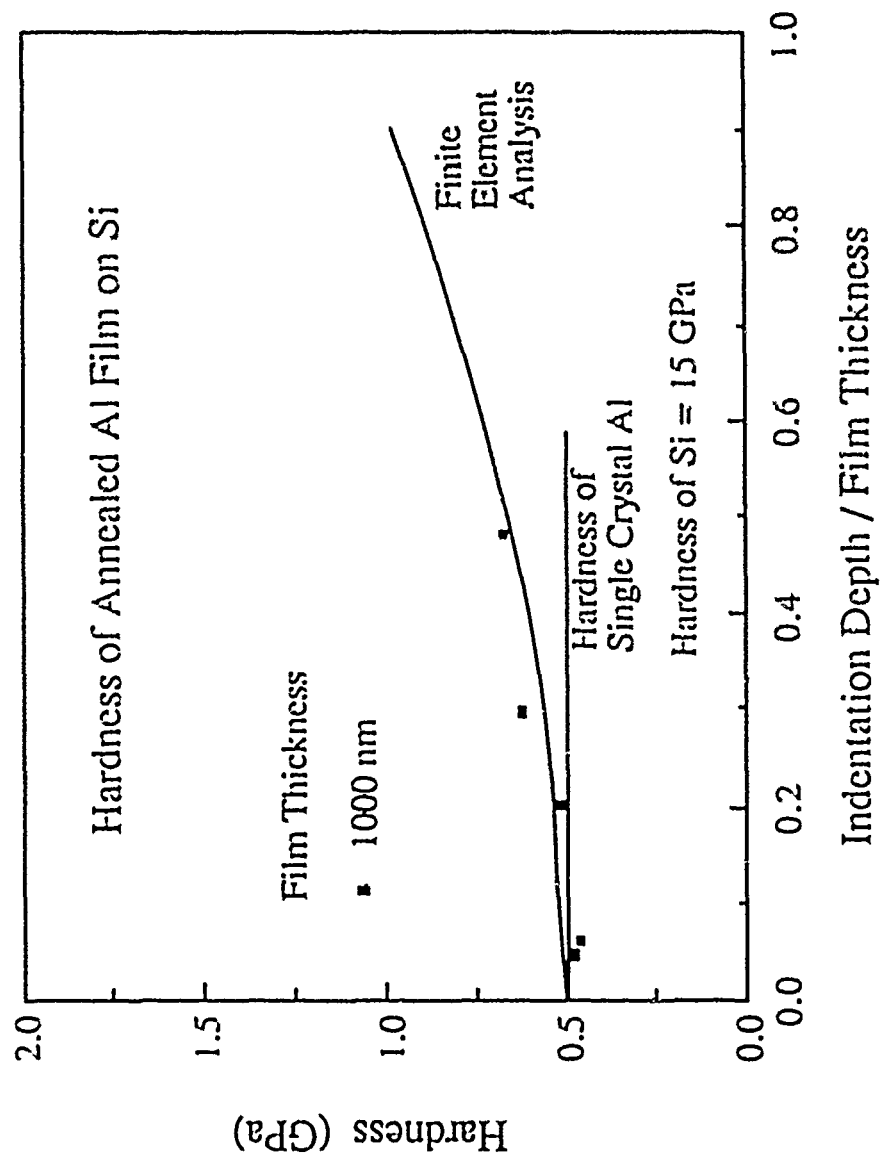
38

Figure 37. Hardness of an Al film on a Si substrate as a function of the depth of indentation [29]. The results of a finite element analysis are also shown for comparison [52]. The continuum treatment gives a good description of the data for a film of this thickness and grain size.
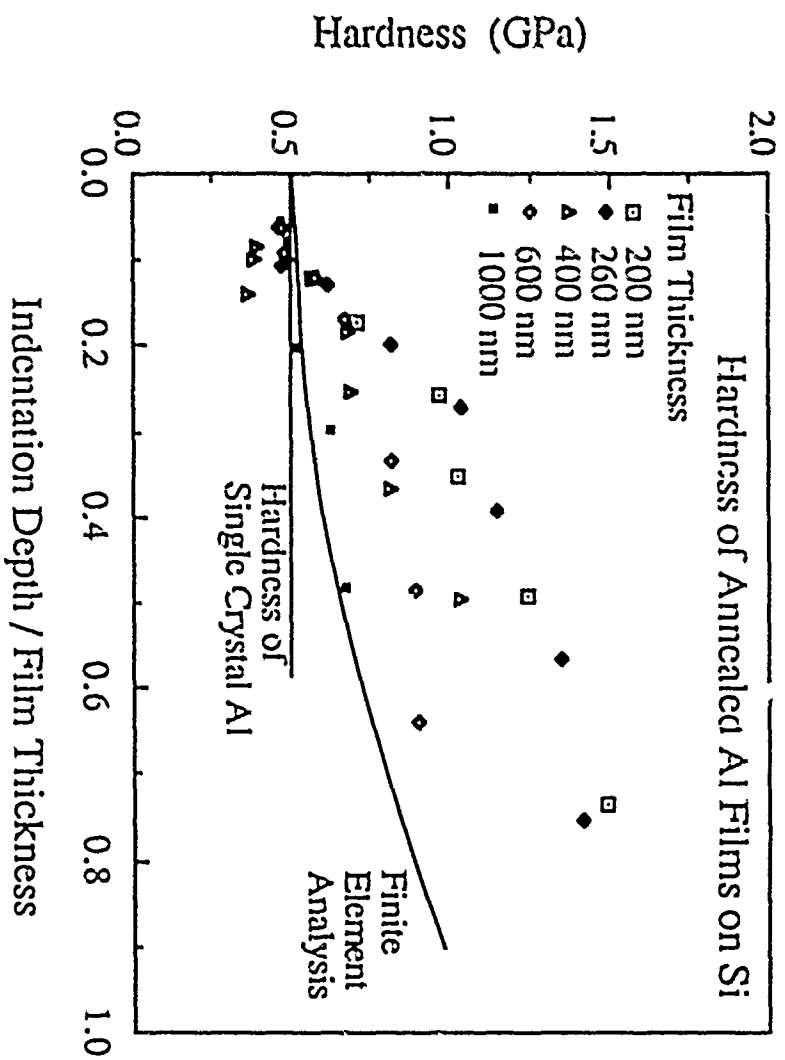
Hardness of Annealed Al Films on Si

Hardness (GPa)

Indentation Depth / Film Thickness

Film Thickness
□ 200 nm
◆ 260 nm
▲ 400 nm
◇ 600 nm
■ 1000 nm

Hardness of
Single Crystal Al

Finite
Element
Analysis

Figure 38. Hardness of a series of Al films on Si substrates as a function of the depth of indentation [29]. The results of a finite element analysis are also shown for comparison [52]. The continuum treatment does not give a good description of the data for the thinner films with the finer grain sizes.

Creep and Time-Dependent Deformation

Because the rate of movement of the indenter can be measured with the Nanoindenter, it is possible to study the effects of strain rate on deformation in the indentation mode. This provides an additional experimental variable in the study of thin film deformation. Here we describe the techniques that have been developed to study strain rate effects in indentation and show results that have been obtained for bulk materials.

As discussed above, the stresses and strains associated with indentation are non-uniform. Thus the stress and strain rate vary with radial distance from the point of contact. As a result, only nominal values of stress and strain rate can be calculated from the measured indentation force and velocity. As shown in Fig. 39, we define the stress as the average pressure acting on the contact surface, $\sigma=P/A$ and the strain rate as the indentation velocity divided by the indentation depth, $\dot{\epsilon}=(1/h)(dh/dt)$. Stresses and strain rates at any point in the sample are expected to scale with these quantities. Thus the strain rate sensitivity can be found using these nominal quantities.

For soft metals and alloys, the elastic displacements associated with indentation are usually small so that the total depth of indentation can be taken to be the plastic depth. For this reason the nominal strain rate can be calculated from the indentation velocity and the total depth at that point in the indentation. A complication does arise, however, for soft metals that exhibit strain hardening. The measured hardness of these materials typically decreases with depth of indentation. This may be caused in part by the fact that the "geometrically-necessary" dislocation density decreases with increasing indentation size. A simple argument reveals that the "geometrically-necessary" dislocation density varies reciprocally with the depth of indentation and this can be used to explain the depth dependence of the hardness. Because of these effects the relationship between stress and strain rate must be determined at a fixed indentation depth to avoid complications associated with strain hardening. Two methods to do this have been developed by Mayo [53-55]. In one technique, a series of indentations is made at various loading rates and a stress and strain rate is determined for each test at a particular depth [53-54]. Figure 40 shows the results of such measurements for coarse-grained samples of Pb and Sn and for a fine-grained sample of a Pb-Sn alloy. The slopes of these curves indicate the stress exponent for creep flow, which is the reciprocal of the strain rate sensitivity. We see that the superplastic Pb Sn alloy has a high strain rate sensitivity of about 0.5, by comparison the coarse-grained samples of Pb and Sn have much lower strain rate sensitivities. These results are in good agreement with tests on bulk samples of these materials. An alternate method for determining the stress exponent (or strain rate sensitivity) has been developed recently and used for the study of strain rate sensitivity of nanophase TiO_2

INDENTATION STRESS AND STRAIN RATE

Indenter

P

h

Contact Area
$A = Kh^2$
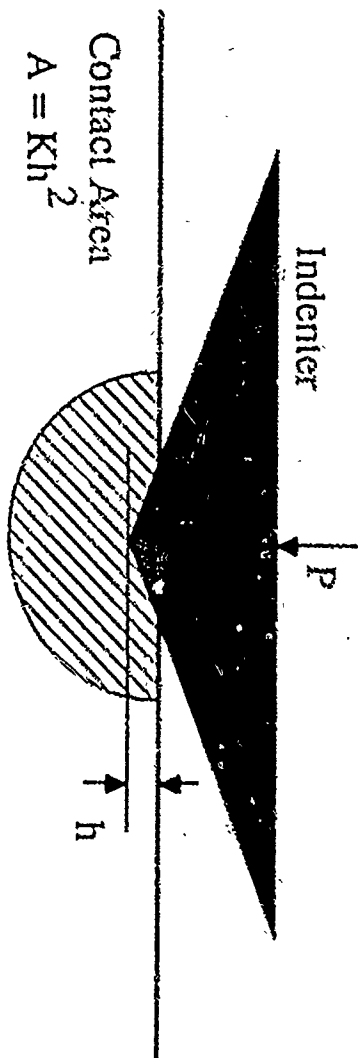
Stress
$\sigma = \dfrac{P}{A}$

Strain Rate
$\dot{\varepsilon} = \dfrac{1}{h}\dfrac{dh}{dt}$

Figure 39.   Definitions of nominal stress and strain rate used in the study of strain rate effects in indentation.
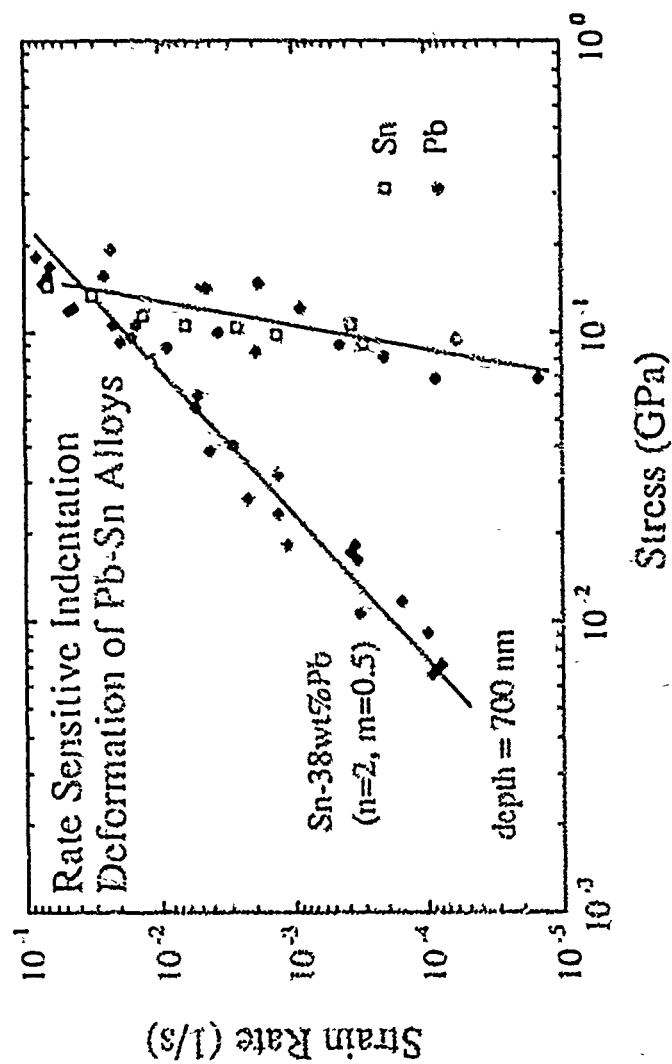
Figure 40.   Stain rate – stress relations for coarse-grained Pb and Sn and for fine-grained (superplastic) samples of a Pb-Sn alloy, obtained through indentation experiments [53-54]. The indentation measurements give a high strain rate sensitivity for fine-grained Pb-Sn as expected. These data were obtained by making various indentations at different loading rates and measuring the stress and strain rate at a particular depth.

[55]. In this technique the rate of movement of the indenter into the material is monitored while the load is held fixed after a rapid initial load application. The nominal stresses and strain rates produced in superplastic Pb-Sn using this technique are shown in Fig. 41. Again we see that a high strain rate sensitivity of 0.5 is correctly found for this material. These methods can also be used to study the strain rate sensitivity of plastic flow in thin films, provided the strain rate sensitivity is manifested at room temperature where the Nanoindenter can be used. Tests of this kind should provide additional information about the controlling mechanisms of deformation in thin films.

VII.   Mechanical Testing of Thin Films Using Microbeam Deflection Techniques

Although the elastic and plastic properties of thin films can be studied most easily using the Nanoindenter in the indentation mode, this mode of testing has the disadvantage that the material being deformed is under a large hydrostatic pressure. In the case of porous materials, this pressure can cause densification to occur and, in other cases, phase transformations can be induced. These complications sometimes make it difficult to study the elastic and plastic properties alone. To avoid the problems associated with indenter pressure, Weihs et al. [56-57] and Hong et al. [58] have developed techniques for creating micromechanical test samples in the shape of singly supported cantilever beams.

Microbeam Fabrication

The beams are typically 1 μm thick, 20 μm wide and 50-100 μm long. Lithographic patterning and anisotropic etching (or chemical micromachining) techniques are used to make these structures. Once fabricated, the beams are tested in the bending mode using the Nanoindenter. Both elastic and plastic deformation of thin film materials can be studied in this way. Because the deformation occurs mainly at the fixed end of the beam, it is not affected by the pressure under the indenter. Figures 42 and 43 show scanning electron micrographs of microbeams of SiO_2 and Au made in this way. The clean faces of the pit in the silicon wafer are produced by anisotropic etching effects. The side faces of the pit are (111) planes with edges along the bottom of the pit that are parallel to the <110> directions. The crystallographic nature of the pit is shown schematically in Fig. 44 for the gold beams.
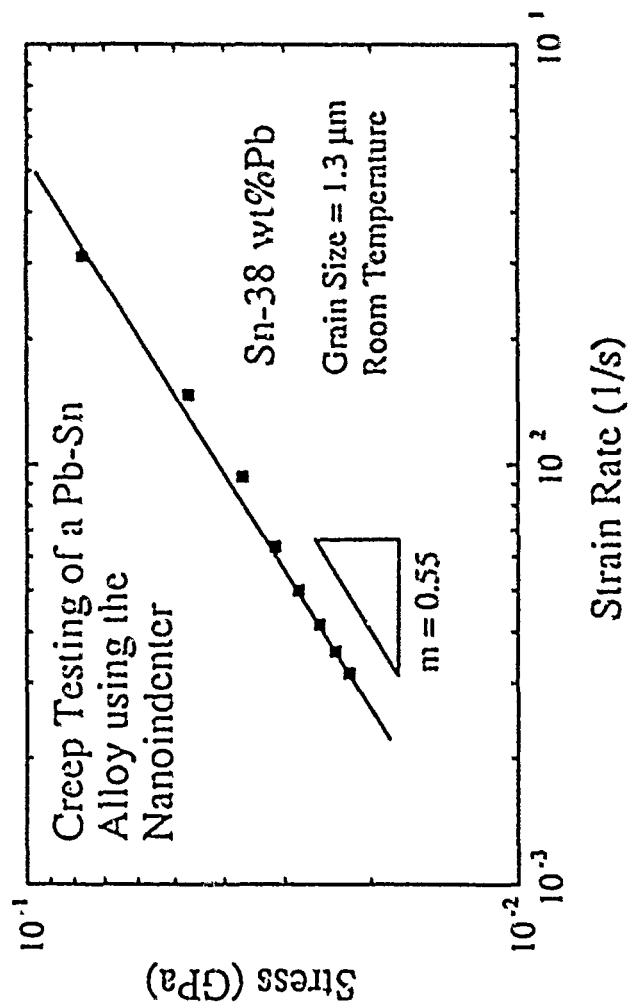
Figure 41. Stress - Strain rate relations obtained for fine-grained Pb-Sn using the Nanoindenter in the creep mode [55]. The data were taken from one indentation experiment in which the load was quickly applied and the stress and strain rate measured as the indenter sinks into the sample.
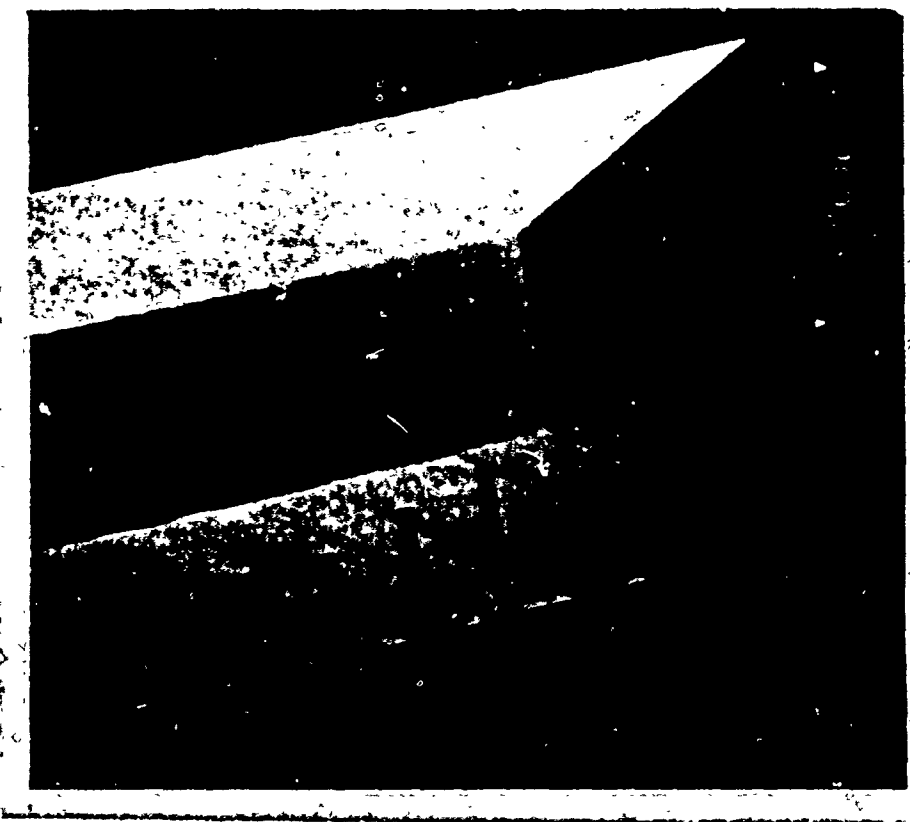
Figure 42. Scanning electron micrograph of cantilever microbeams of thermal SiO_2 made by integrated processing techniques [56-58]. The beams extend over a pit in the underlying Si substrate.
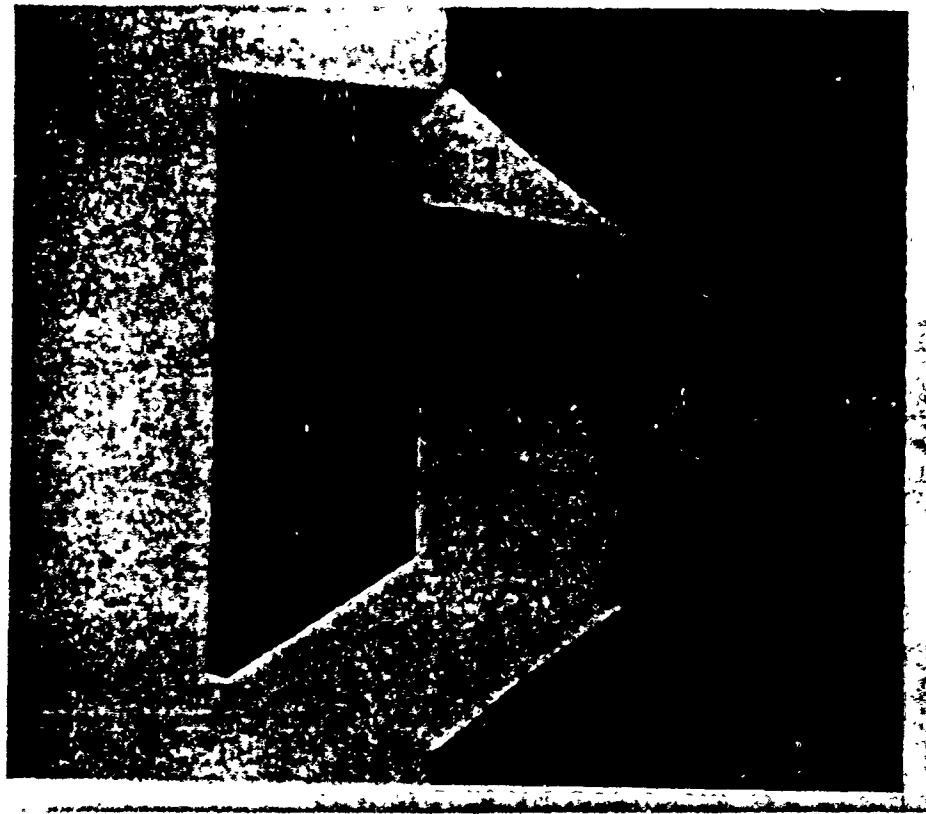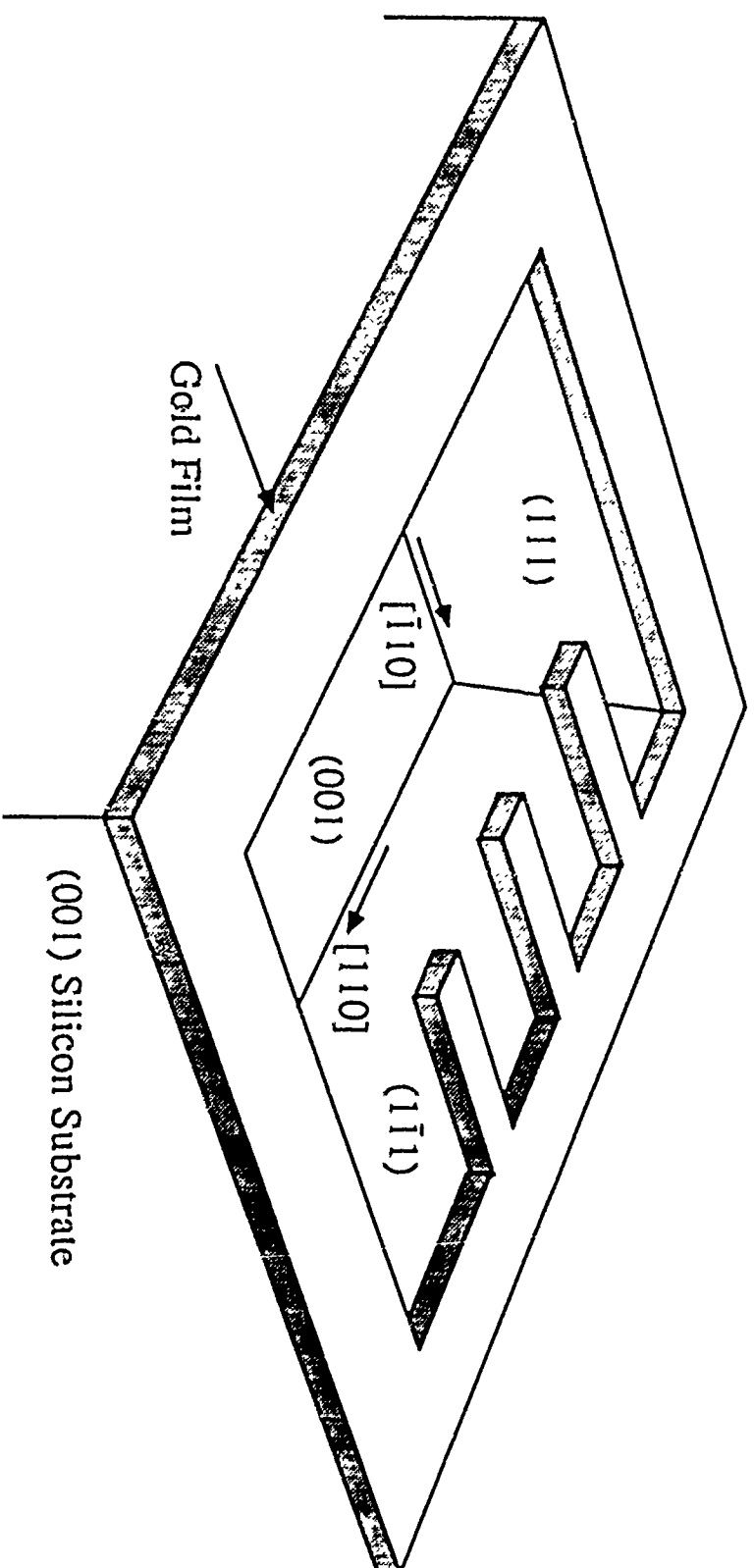
Figure 43. Scanning electron micrograph of cantilever microbeams of Au made by integrated processing techniques [56-58]. The beams extend over a pit in the underlying Si substrate.

Microbeam Test Samples

Gold Film

(001) Silicon Substrate

(111)

[īĪ0]

(001)

[Ī10]

(1Ī1)

Figure 44. Schematic diagram of microbeams of Au extending over an etch pit in the underlying Si substrate. The crystallographic orientation of pit is created by anisotropic etching of Si.

Microbeam Testing

The testing of the microbeams is accomplished by deflecting the beams with the Nanoindenter, as shown in Fig. 45. The indenter is first positioned over one point on the microbeam and then moved toward the beam by gradually increasing the force. Before the indenter contacts the beam, the imposed force is supported entirely by the suspending springs of the Nanoindenter and the relationship between the force and displacement in this regime is determined by the stiffness of these springs. The stiffness of the system increases as soon as the indenter makes contact with the beam. This is shown in Fig. 46 for a test on a microbeam of SiO_2. In order to obtain the bending properties of the beam, it is necessary to subtract the forces associated with the deflection of the suspending springs and to measure displacement from the point at which the indenter first makes contact with the beam. When this is done for microbeams of SiO_2, the force-displacement relations shown in Fig. 47 are found.

Elastic Properties

The slopes of the lines in Fig. 47 give the stiffnesses of the beams and from these measurements the elastic modulus of the beam material can be obtained, using appropriate beam bending relations and the dimensions of the beams. The elastic properties of several thin film materials found using the beam bending technique are shown in Table 1 and are compared with elastic properties obtained from indentation measurements. The beam bending results are in better agreement with values taken from the literature than are the indentation results.

Plastic Properties

The yield strengths of thin metal films can also be measured using the beam bending technique. Beam bending force-displacement relations are shown in Fig. 48 for microbeams of gold. The curve that starts at the origin represents the force-displacement relation associated with an initial loading and unloading of a beam. The loading curve is approximately linear, except at the highest loads where some non-elastic behavior is noted. We also see that the displacement does not return to zero when the load is completely removed. This indicates that plastic deformation occurred during the test, most likely near the substrate support where the bending stresses in the beam are maximum. A subsequent loading experiment is also shown in the figure, for clarity the second curve has been displaced arbitrarily along the displacement axis. We note that the loading portion of the second curve remains linear up to about 60 μN, the maximum load reached in the first loading cycle. Non-linear, plastic behavior is observed only at higher loads. This indicates
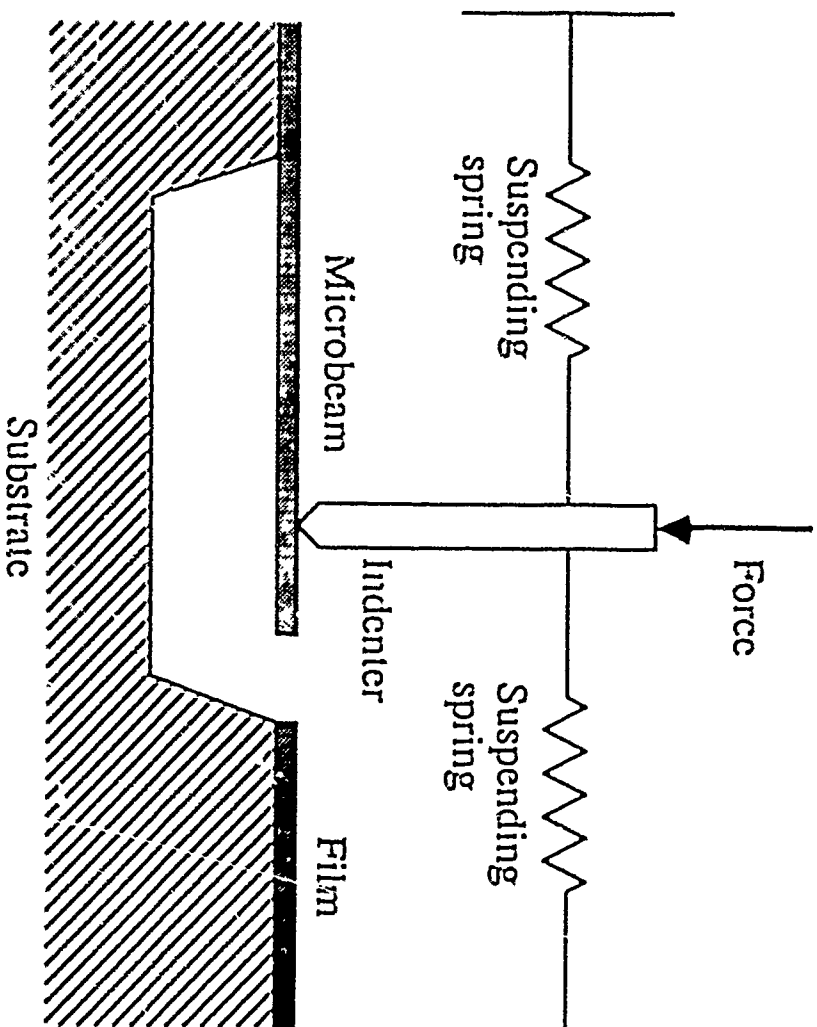
MECHANICAL TESTING OF MICROBEAMS

Figure 45. Schematic diagram of the microbeam testing procedure. The indenter is placed over the beam and gradually lowered until contact with the beam is made. After contact with the beam is made, the imposed force is supported by both the suspending springs and the deflection of the microbeam.
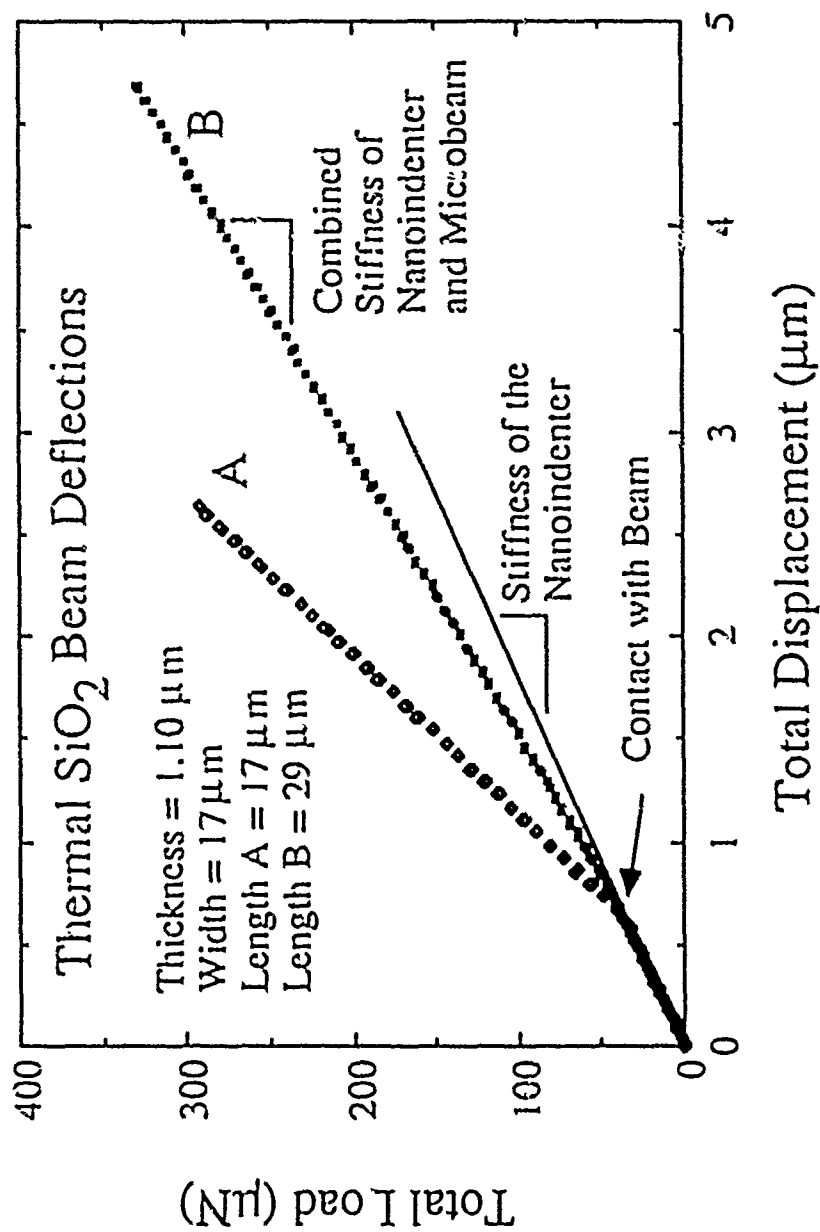
Figure 46. Total applied load and total displacement measured for two deflections of a thermal SiO$_2$ cantilever microbeam [56].
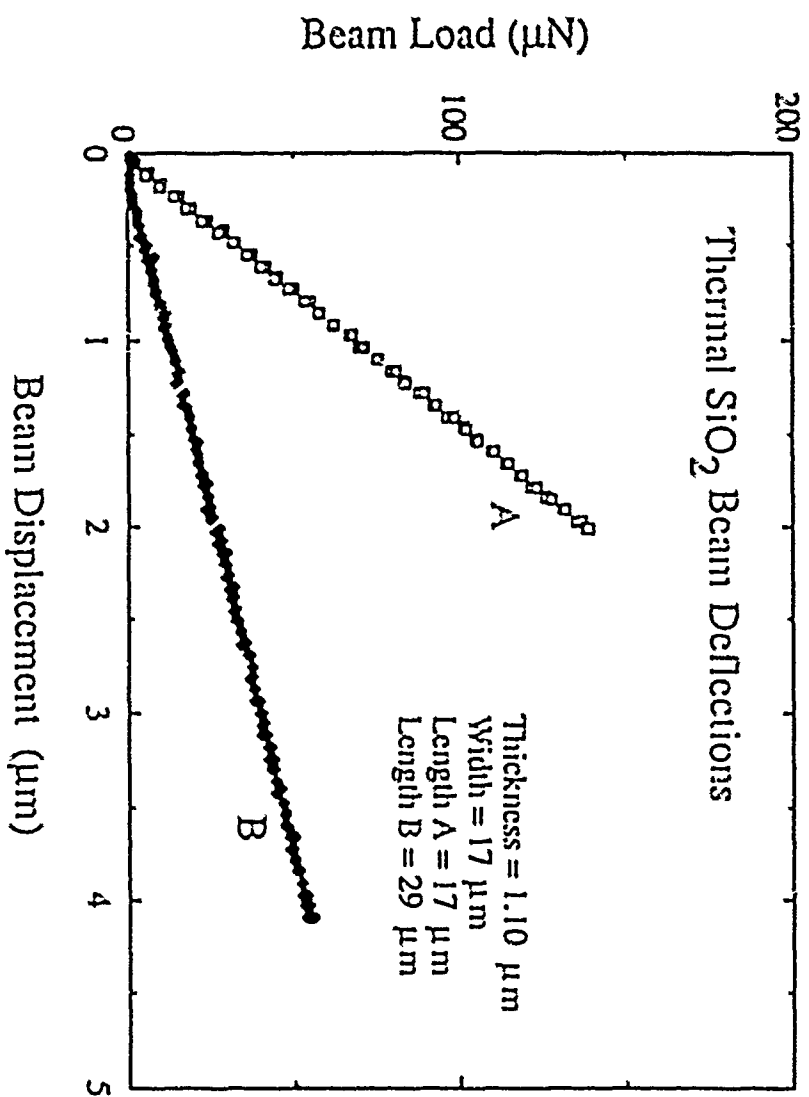
Figure 47. Net load and net displacement measured for two deflections of a thermal SiO$_2$ cantilever microbeam [56]. The data shown here were taken from Fig. 46.
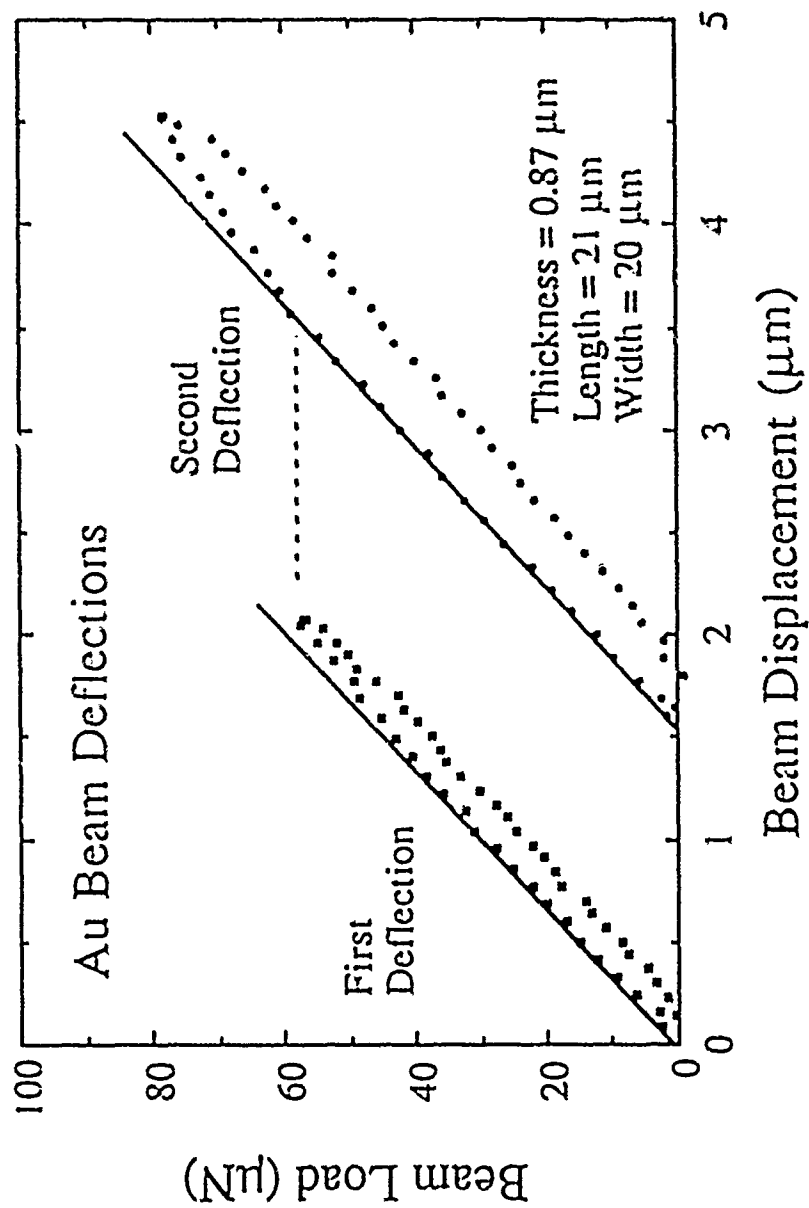
Figure 48. Net load and net displacement measured for cantilever microbeams of Au [56]. The data from two different different deflections are shown. A small amount of plastic flow occurs during the first deflection and the beam is strain hardened. A high load is required to cause yielding during the second deflection.

that the beam was strain hardened during the first loading cycle. The increased amount of plastic deformation that occurs during the second loading cycle is indicated by the greater offset observed when the load is reduced to zero. The yield strengths and the strain hardening properties of metal beams can be found by calculating the stresses in the beam during the course of non-elastic bending. The yield strength of thin film gold obtained in this way is shown in Table 1, and can be compared with the hardness of the same material obtained using indentation. We note that the yield strength of gold measured by beam bending is about 1/3 of the hardness measured by indentation, a result consistent with the classical plasticity relation, $H=3\sigma_y$.

<u>Table 1</u>

Elastic and Plastic Properties Measured for Thin Film Coatings

	Beam Test		Indentation Test		Literature
	E	σ_y	E	H	E
Si	163	>3.39	188	13	169
Thermal SiO2	64	>0.57	85	12	57,66
LTO	44	-	83	10	-
Au	57	0.34	74	1.0	39-78, 80[**]

* All data in GPa

** First range is from measurements on thin films, second value is for bulk gold. Gold is electron beam deposited.

that the beam was strain hardened during the first loading cycle. The increased amount of plastic deformation that occurs during the second loading cycle is indicated by the greater offset observed when the load is reduced to zero. The yield strengths and the strain hardening properties of metal beams can be found by calculating the stresses in the beam during the course of non-elastic bending. The yield strength of thin film gold obtained in this way is shown in Table 1, and can be compared with the hardness of the same material obtained using indentation. We note that the yield strength of gold measured by beam bending is about 1/3 of the hardness measured by indentation, a result consistent with the classical plasticity relation, $H=3\sigma_y$.

<u>Table 1</u>

Elastic and Plastic Properties Measured for Thin Film Coatings

| | Beam Test | | Indentation Test | | Literature |
	E	σ_y	E	H	E
Si	163	>3.39	188	13	169
Thermal SiO2	64	>0.57	85	12	57,65
LTO	44	-	83	10	-
Au	57	0.34	74	1.0	39-78, 80[**]

* All data in GPa

** First range is from measurements on thin films, second value is for bulk gold. Gold is electron beam deposited.

VIII. Concluding Remarks

We have described the stresses in thin films on substrates, as well as some of the mechanisms by which thin film deformation occurs. This study leads to an understanding of the formation of misfit dislocations in heteroepitaxial structures as well the high strengths of thin metal films on substrates. Because non-traditional testing techniques are required, much of the attention of this paper has been focused on these techniques. We have shown that laser scanning techniques for measuring substrate curvature provide a convenient way to study the biaxial stresses in thin films and the deformation processes that occur at high temperatures. Sub-micron indentation testing techniques for studying the hardness, elastic modulus and time-dependent deformation properties of thin films on substrates have also been described. We have shown that the elastic and plastic properties of thin films can be studied by deflecting cantilever microbeams made by integrated circuit manufacturing methods.

The motivation for much of the work described here has been the desire to understand the mechanical properties of the thin film materials used in microelectronic and magnetic devices. As such, the work may seem to be a move away from the traditional field of mechanical properties of structural, load bearing, materials. While the experimental techniques described here may appear to be specific to the problems that arise in microelectronic and magnetic applications, they are equally useful in the study of bulk structural materials. It is important to remember that high performance structural materials, like electronic materials, are also finely structured. There is a need to obtain information about the mechanical properties of individual features in the microstructures of structural materials. The techniques that are available for the study of microelectronic and magnetic thin films can also be used to study the mechanical properties of some of these microstructural features. In this way, the study of mechanical properties of thin films is complimentary to the study of bulk structural materials. Thus, the mechanical properties of thin films represents but another chapter in the long history of the study of the relationship between microstructure and mechanical properties in the field of metallurgy and materials science.

.

Acknowledgements

References

1. G. E. Henein and W. R. Wagner, J. Appl. Phys., 54, 6395-6400 (1983).

2. K. Roll, J. Appl. Phys., 47, 3224-3229 (1976).

3. P. H. Townsend, D. M. Barnett and T. A. Brunner, J. Appl. Phys., 62, 4438-4444 (1987).

4. J.H. Van der Merwe, J. Appl. Phys., 34, 123-127 (1963).

5. J.W. Matthews and A.E. Blakeslee, J. Cryst. Growth, 27, 118-125 (1974).

6. J.W. Matthews and A.E. Blakeslee, J. Cryst. Growth, 29, 273-280 (1975).

7. J.W. Matthews, J. Vac. Sci. Technology, 12, 126-133 (1975).

8. E. Kaspar and H.-J. Herzog, Thin Solid Films, 44, 357-370 (1977).

9. J.C. Bean, L.C. Feldman, A.T. Fiory, S. Nakahara and I.K. Robinson, J. Vac. Sci. Technology A, 2, 436-440 (1984).

10. J.Y. Tsao, B.W. Dodson, S.T. Picraux and D.M. Cornelison, Phys. Rev. Lett., 59, 2455-2458 (1987).

11. C. Gronet, PhD Dissertation, Stanford University (1988).

12. Y. Kohama, Y. Fukuda and M. Seki, Appl. Phys Lett, 52, 380-382 (1988).

13. P.L. Gourley, I.J. Fritz and L.R. Dawson, Appl. Phys. Lett., 52, 377-379 (1988).

14. P.M.J. Maree, J.C. Barbour, J.F. van der Veen, K.L. Kavanagh, C.W. T. Bulle-Lieuwma and M.P.A. Viegers, J. Appl. Phys., 62, 4413-4420 (1987).

15. R. Hull, J.C. Bean, D.J. Werder and R.E. Leibenguth, Appl. Phys. Lett., 52, 1605-1607 (1988).

16. B.W. Dodson and J.Y. Tsao, Appl. Phys. Lett., 51, 1325-1327 (1987).

17. L.B. Freund, J. Appl. Mech., 54, 553-557 (1987).

18. D.M. Barnett, private communication

19. P. Hassen and H. Alexander, Solid State Physics, 22, 27-158 (1968).

20. H. Steinhardt and S. Schafer, Acta Metall., 19, 65-70 (1971).

21. H. Steinhardt and P. Haasen, phys. status solidi (a), 49, 93-101 (1978).

22. W. Hagen and H. Strunk, Appl. Phys., 17, 85-87 (1978).

23. K. Rajan and M.Denhoff, J. Appl. Phys., 62, 1710-1716 (1987).

45

24. P.A. Flinn, D.S. Gardner and W.D. Nix, IEEE Trans. on Electron Devices, ED-34, 689-699 (1987).

25. M. F. Doerner and S. Brennan, J. Appl. Phys., 63, 126-131 (1988).

26. P. H. Townsend, Ph. D. Dissertation, Stanford University (1987).

27. D. S. Gardner, T. L. Michalka, P. A. Flinn, T. W. Barbee Jr., K. C. Saraswat and J. D. Meindl, Proc. 2nd International IEEE VLSI Multilevel Interconnection Conference, (1985) p. 102-110.

28. T.S. Kuan and M. Murakami, Metall. Trans. A, 13A, 383-391 (1982).

29. M. F. Doerner, D. S. Gardner and W. D. Nix, J. Mater. Res., 1, 845-851 (1986).

30. M.F. Doerner, Ph.D. Dissertation, Stanford University (1987).

31. N. Hansen, Acta Metall., 25, 863-869 (1977).

32. R.W. Armstrong, in Advances in Materials Research, 4, H. Herman, Ed., Intescience, New York (1970) p. 67.

33. A. J. Griffin, Jr., F. R. Brotzen and C. Dunn, Scripta Metall.; 20, 1271-1272 (1986).

34. A. J. Griffin, Jr., F. R. Brotzen and C. Dunn, Thin Solid Films, 150, 237-244 (1987).

35. M. Nishibori and K. Kinosita, Thin Solid Films, 48, 325-331 (1978).

36. D. Newey, M. A. Wilkins and H. M. Pollock, J. Phys. E, 15, 119-122 (1982).

37. J. Pethica, R. Hutchings and W. C. Oliver, Phil. Mag., A48, 593-606 (1983).

38. J. L. Loubet, J. M. Georges, J. M. Marchesini and G. Meille, J. Tribology, 106, 43-48 (1984).

39. P.E. Wierenga and A.J.J. Franken, Philips Tech Rev., 42, 85-92 (1985).

40. H. Bangert, A. Kaminitschek, A. Wagendristel, A. Barna, P.B. Barna and G. Radnoczi, Thin Solid Films, 137, 193-198 (1986).

41. S-P. Hannula, D. Stone and C.-Y. Li, Mater. Res. Symp. Proc., 40, 217-224 (1985).

42. D. Stone, W.R. LaFontaine, P. Alexopoulous, T.-W. Wu and C.-Y. Li, J. Mater. Res., 3, 141-147 (1988).

43. J.B. Pethica and W.C. Oliver, Mater. Res. Symp. Proc., 130, 13-23 (1989).

44. D. Stone, W. La Fontaine, S. Ruoff and C. Y. Li, Mater. Res. Soc. Proc., 72, 43-49 (1986).

45. H.-Y. Yu and J.C.M. Li, J. Mater. Sci., 12, 2214-2222 (1977).

46. H.-Y. Yu, M.A. Imam and B.B. Rath, J. Mater. Sci., 20, 636-642 (1985).

47. H.-Y. Yu, S.C. Sanday and B.B. Rath, Naval Research Laboratory Report, No. 9168, Jan. 12 (1989).

48. M. F. Doerner and W. D. Nix, J. Mater. Res., 1, 601-609 (1986).

49. I. N. Sneddon, Int. J. Eng. Sci., 3, 47-62 (1965).

50. R.B. King, Int. J. Solids and Structures, 23, 1657-1664 (1987).

51. T.F. Page, W.C. Oliver and C.J. McHargue, to be published in J. Mater. Sci.

52. A.K. Bhattacharya and W.D. Nix, Int. J. Solids and Structures, 24, 1287-1298 (1988).

53. M.J. Mayo and W.D. Nix, Acta Metall., 36, 2183-2192 (1988).

54. M.J. Mayo and W.D. Nix, Proc. 8th International Conference on Strength of Metals and Alloys, Tampere, Finland, Eds: P.O. Kettunen, T.K. Lepisto and M.E. Lehtonen, Pergamon Press (1988) p.1415-1420.

55. M.J. Mayo, R.W. Siegel, A. Narayasamy and W.D. Nix, submitted to J. Mater. Research.

56. T.P. Weihs, S. Hong, J.C. Bravman and W.D. Nix, J. Materials Research, 3, 931-942 (1988).

57. T.P. Weihs, S. Hong, J.C. Bravman and W.D. Nix, Mater. Res. Symp. Proc., 130, 87-92 (1989).

58. S. Hong, T.P. Weihs, J.C. Bravman and W.D. Nix, Mater. Res. Symp. Proc., 130, 93-98 (1989).

DECOHESION OF A CUT PRESTRESSED FILM ON A SUBSTRATE

Henrik M. Jensen

John W. Hutchinson

Kyung-Suk Kim

DECOHESION OF A CUT PRESTRESSED FILM ON A SUBSTRATE[*]

Henrik M. Jensen
Department of Solid Mechanics
The Technical University of Denmark
DK-2800 Lyngby, Denmark

John W. Hutchinson
Division of Applied Sciences
Harvard University
Cambridge, MA 02138

Kyung-Suk Kim
Department of Theoretical and Applied Mechanics
University of Illinois at Urbana-Champaign
Urbana, IL 61801

Abstract

Thin films or coatings on a substrate are often under a state of residual biaxial tension. This paper analyzes the cut test wherein a straight cut which is long compared to the film thickness is made through the film down to the substrate. Depending on the elastic properties of the materials, the residual stress level, and the toughness of the interface bond between the film and substrate, the cut may induce extensive decohesion, limited decohesion or no decohesion at all. The main thrust of this paper is concerned with the shape and extent of the zone of decohesion when extensive decohesion occurs. The interface crack front at the boundary of the decohered region is subject to mixed mode conditions involving modes 1, 2 and 3. The shape of the decohesion region depends sensitively on the way in which the mode 3 contribution enters the decohesion criterion. Observations of shapes for a polyimide film on a glass substrate permit the relative importance of the mode 3 contribution to be inferred when the decohered film makes no contact with the substrate. Some experimental observations indicate that the decohesion process is quite complex when the decohered film makes contact with the substrate. A prototypical criterion for mixed mode interfacial fracture is proposed for cases in which the interfacial crack remains open.

---

[*] Dedicated to the memory of Charles D. Babcock

1. INTRODUCTION

A thin film or coating bonded to a substrate and under a state of residual biaxial tension will decohere starting from a flaw at a free edge or from a through-cut if the residual stress is sufficiently high and/or the bond is sufficiently weak (Evans, Drory and Hu, 1988). If the substrate has low toughness compared to the interface, cracking may occur in the substrate (Evans, Drory and Hu, 1988, and Suo and Hutchinson, 1989), but that possibility is not considered here. Suppose the film is in a state of equal biaxial tension σ and suppose the film thickness t is very small compared to the substrate thickness so that the substrate can be considered to be infinitely thick. Let E, μ and ν be Young's modulus, shear modulus and Poisson's ratio of the film and E_s, μ_s and ν_s be the corresponding quantities for the substrate. The energy release rate, $\mathcal{G}$, of a two-dimensional plane strain crack on the interface and emerging from a through-cut is shown in Fig. 1. This result was obtained by Thouless, Cao and Mataga (1989) using finite element techniques for the case where the elastic properties of the substrate coincide with those of the film. When the interface crack is longer than 3 to 5 film thicknesses it approaches steady-state conditions with a length-independent energy release rate given by

$$\mathcal{G} = \frac{1}{2}\frac{1-\nu^2}{E}\sigma^2 t \tag{1}$$

This steady-state result is independent of the elastic properties of the substrate.

The plane strain interface crack is characterized by two stress intensity factors, K_1 and K_2, such that on the interface a distance r ahead of the tip the normal and shear stresses are given by

$$(\sigma_{22},\sigma_{12}) = (K_1,K_2)(2\pi r)^{-1/2} \tag{2}$$

The two Dundurs parameters measuring the elastic mismatch between the film and the substrate are

$$\alpha = \frac{E/(1-\nu^2) - E_s/(1-\nu_s^2)}{E/(1-\nu^2) + E_s/(1-\nu_s^2)} \quad , \quad \beta = \frac{1}{2}\frac{\mu(1-2\nu_s) - \mu_s(1-2\nu)}{\mu(1-\nu_s) + \mu_s(1-\nu)} \tag{3}$$

In writing (2) it has been assumed that $\beta = 0$ (either exactly or as an approximation), and thus we are sidestepping some of the difficulties associated with the more pathological oscillatory interface singularity. Throughout the paper we will take $\beta = 0$. Under this circumstance and under combined mode 1 and 2,

$$\mathfrak{I} = \frac{1}{2}\left(\frac{1-v^2}{E} + \frac{1-v_i^2}{E_i}\right)(K_1^2 + K_2^2) \tag{4}$$

assuming $K_1 > 0$ so that the crack is open. The steady-state plane strain problem for decohesion driven by a tensile stress σ has $K_1 > 0$ but is heavily mixed mode with the phase of the stress intensity factors, $\psi = \tan^{-1}(K_2/K_1)$, depending on α as shown in Fig. 2 (Hutchinson (1989) and Suo and Hutchinson (1988)).

In general, interface toughness is strongly mode-dependent (Cao and Evans (1988)), usually with the energy release rate needed to drive the crack increasing with increasing proportion of mode 2 to mode 1. The condition for decohesion of a plane strain crack can be written as

$$\mathfrak{I} = \mathfrak{I}_c(\psi) \tag{5}$$

where $\mathfrak{I}_c(\psi)$ is the mode-dependent interface toughness which must be determined by test. The critical stress associated with steady-state plane strain decohesion from (1) and (5) is

$$\sigma_c = \left(\frac{2E\mathfrak{I}_c(\psi)}{(1-v^2)t}\right)^{1/2} \tag{6}$$

where ψ is found from Fig. 2.

The general criterion (5) includes two limiting cases which will be considered later. An "ideally brittle" interface is defined as an interface with a mode-independent toughness, i.e.

$$\mathfrak{I}_c(\psi) = \mathfrak{I}_c^0 \tag{7}$$

One imagines that the total energy released goes into creation of the new surfaces. At the other extreme, consider an interface crack where the tip is fully shielded from any effect of K_2; that is, assume the condition for crack advance is $K_1 = K_1^c$, independent of K_2. The fully shear-shielded criterion is

$$\mathcal{S}_c = \mathcal{S}_c^0 / \cos^2 \psi \tag{8}$$

where $\mathcal{S}_c^0$ is the pure mode 1 toughness related to K_1^c by (4). Experimental data for an epoxy/glass interface obtained by Cao and Evans (1988) over a wide range of modes 1 and 2 fell somewhat below the trend of (8) and well above (7). Using micromechanical modeling, Evans and Hutchinson (1989) produced a family of criteria for combined mode 1 and 2 interface fracture which depended on a single nondimensional parameter characterizing the roughness of the failed interface. Their family of criteria reduced to (7) for a perfectly smooth interface and to (8) in the limit of a very rough interface.

In the cut test, a straight cut of length $2L$, which is very long compared to the film thickness, is made through the film to the substrate. If consideration is limited to purely elastic behavior in the film and substrate and if substrate cracking does not occur, there are three response regimes as depicted in Fig. 3. Let σ_i be the lowest value of σ needed to initiate decohesion at the intersection of the through-cut and the interface. Usually, σ_i will be less than σ_c (c.f. Fig. 1), although, in general, the size of σ_i relative to σ_c will depend on α and on the ψ-dependence of $\mathcal{S}_c(\psi)$. If $\sigma_i < \sigma_c$, as assumed in Fig. 3, then no decohesion can occur if

$$\sigma < \sigma_i \tag{9}$$

If

$$\sigma_i < \sigma < \sigma_c \tag{10}$$

the extent of the decohered region from the cut in its central portion will be determined by a plane strain relation such as that in Fig. 1 and will be limited to distances perpendicular to the cut on the order of, typically, 1 to 5 times t. If

$$\sigma > \sigma_c \tag{11}$$

there will be extensive decohesion spreading out over an area controlled by the length of the cut. For material combinations for which $\sigma_i > \sigma_c$, there are two response regimes: $\sigma < \sigma_i$ with no decohesion and $\sigma > \sigma_i$ with extensive decohesion.

2. MODEL OF THE CUT TEST ($\sigma > \sigma_c$)

The regime of extensive decohesion is considered with $\sigma > \sigma_c$ (and $\sigma > \sigma_i$ if $\sigma_i > \sigma_c$).
If $L \gg t$, it can be anticipated that the extent of the decohesion perpendicular to the cut will be
determined by L and will be very large compared to t. For one experimental system discussed in
Section 4, the lateral extent of the decohered region is about one hundred times t. The stress
distribution in the cut decohered film is analyzed as a two-dimensional plane stress problem.
Assume for the moment that the boundary C of the decohered region is known. As depicted in
Fig. 4, the stress distribution within the decohered region of the cut film is the sum of the stresses
in two plane stress problems, (i) and (ii). Problem (i) is the trivial uniform equal biaxial stress
state σ and Problem (ii) is the plane stress distribution for a film clamped along C and subject to
a normal traction σ along the cut, i.e.

$$u_1 = u_2 = 0 \qquad \text{on } C$$

$$\sigma_{12} = 0 \text{ and } \sigma_{22} = -\sigma \qquad \text{on } x_2 = 0, |x_1| < L$$

(12)

The boundary C of the decohered region has been depicted as standing off the ends of the
cut in Fig. 4. Whether the boundary stands off or is attached to the ends of the cut turns out to be a
significant issue which is tied to the details of the interface decohesion criterion. This issue will be
dealt with in Section 3.

The stresses in the film resolve into a normal component, σ_{nn}, and a tangential component,
σ_{nt}, to the boundary C as shown in Fig. 5. The steady-state energy release rate for a *straight*
crack front when the film has *uniform* normal and tangential stresses is, instead of (1),

$$\mathfrak{I} = \frac{1}{2} \frac{(1-v^2)}{E} \Delta\sigma_{nn}^2 t + \frac{1}{2} \frac{1}{\mu} \Delta\sigma_{nt}^2 t$$

(13)

where $\Delta\sigma_{nn}$ and $\Delta\sigma_{nt}$ are the stress *changes* from the state of biaxial tension σ. Equation (13)
holds locally along the crack front if the radius of curvature of C and the length scale over which
$\Delta\sigma_{nn}$ and $\Delta\sigma_{nt}$ vary are large compared to the film thickness t. This is assumed to be the case

in the present model, and the stress changes $\Delta\sigma_{\alpha\alpha}$ and $\Delta\sigma_{\alpha t}$ are calculated from the stresses in Problem (ii).

Locally along the crack front C, $\Delta\sigma_{\alpha\alpha}$ induces mode 1 and mode 2 with K_2 and K_1 in *fixed* proportion characterized by ψ in Fig. 2, while $\Delta\sigma_{\alpha t}$ induces mode 3. Under the three modes of loading the energy release rate is related to the stress intensity factors by (with $\beta = 0$)

$$\mathfrak{S} = \frac{1}{2}\left(\frac{1-v^2}{E} + \frac{1-v_s^2}{E_s}\right)(K_1^2 + K_2^2) + \frac{1}{4}\left(\frac{1}{\mu} + \frac{1}{\mu_s}\right)K_3^2 \tag{14}$$

Thus, by comparing (13) and (14), one has

$$\left(\frac{1-v^2}{E} + \frac{1-v_s^2}{E_s}\right)(K_1^2 + K_2^2) = \left(\frac{1-v^2}{E}\right)\Delta\sigma_{\alpha\alpha}^2 t \tag{15}$$

and

$$\frac{1}{2}\left(\frac{1}{\mu} + \frac{1}{\mu_s}\right)K_3^2 = \frac{1}{\mu}\Delta\sigma_{\alpha t}^2 t \tag{16}$$

where $\psi = \tan^{-1}(K_2/K_1)$ is given by Fig. 2. The above assumes $K_1 > 0$, which holds if $\Delta\sigma_{\alpha\alpha} < 0$. If $\Delta\sigma_{\alpha\alpha} \geq 0$, the crack front does not open, $K_1 = 0$, and (14) and (15) apply with $K_1 = 0$ if the fractured interface is perfectly smooth with no friction.

We now propose a family of criteria governing the advance of the interface crack between the film and the substrate. Since the relative proportion of K_2 to K_1 is fixed all along the front, the shape of the decohered region is not dependent on how ψ enters into the decohesion criterion. Put another way, the observed shape of a decohered region cannot be used to discriminate how the relative proportions of K_2 to K_1 enter the decohesion criterion. As will be seen, however, the shape is strongly dependent on the way in which the mode 3 intensity factor is included in the criterion.

Noting (13), consider the family of criteria specified by

$$\frac{1}{2}\frac{(1-v^2)}{E}\Delta\sigma_{\alpha\alpha}^2 t + \lambda\frac{1}{2}\frac{1}{\mu}\Delta\sigma_{\alpha t}^2 t = \mathfrak{S}_c^* \tag{17}$$

where $\mathcal{G}_c^*$ can be thought of as the critical energy release rate for the plane strain crack in (5) when $\Delta\sigma_{nt} = 0$, i.e. $\mathcal{G}_c^* = \mathcal{G}_c(\psi)$. The parameter λ in (17) is used to adjust the influence of the mode 3 contribution. When $\lambda = 0$, mode 3 has no effect on the condition for crack advance -- the crack tip is fully shielded from any influence of mode 3. At the other limit, $\lambda = 1$, there is no shielding of the crack tip from mode 3. This family of criteria includes the "ideally brittle" interface if $\lambda = 1$ and $\mathcal{G}_c^*$ is identified with $\mathcal{G}_c^0$ in (7). The limit of an interface fully shielded from both mode 2 and mode 3 is also included in (17) if $\lambda = 0$ and $\mathcal{G}_c^*$ is identified with $\mathcal{G}_c(\psi)$ in (8).

When the decohered region remains attached to the ends of the cut, the computed distribution of $\Delta\sigma_{nn}$ along C reported in Section 4 is everywhere negative giving $K_1 > 0$ along the front. When, on the other hand, the decohered region stands off the ends of the cut it is found that there is a small portion of C just opposite the ends where $\Delta\sigma_{nn}$ turns out to be positive implying that the crack is closed along this portion of its front. The steady-state energy release rate (13) still applies when $\Delta\sigma_{nn} > 0$, assuming friction effects are not important, and a criterion in the form of (17) remains plausible. However, one should not necessarily expect the same values of the parameters λ and $\mathcal{G}_c^*$ to hold for both $\Delta\sigma_{nn} < 0$ and $\Delta\sigma_{nn} > 0$. Nevertheless, in the present exploratory study we have used (17) independent of the sign of $\Delta\sigma_{nn}$. The limitations of this assumption for the case in which the decohered region stands off the end of the cut will be discussed in Section 5 along with some experimental evidence which suggests that an adequate criterion is probably lacking when contact between the decohered film and the substrate takes place.

In Appendix I, we propose a class of criteria for characterizing interfacial fracture under combined three-mode conditions, of which (17) is a special case. In Appendix I, the general class of criteria, including the special family (17), are expressed as

$$\mathcal{G} = \mathcal{G}_c(\psi,\phi) \tag{18}$$

where ψ and ϕ are Euler angles in a space of (K_1, K_2, K_3). In the present study, (17) is the convenient form for the decohesion criterion.

We now summarize the model for predicting the shape of the decohered region when $\sigma > \sigma_c$, where in the new notation used in (17)

$$\sigma_c = \left(\frac{2E\mathcal{S}_c^*}{(1-\nu^2)t}\right)^{1/2} \tag{19}$$

The plane stress boundary value problem for determining the boundary C of the decohered region is Problem (ii) of Fig. 4 with boundary conditions (12). The boundary C must be found such that (17) is satisfied everywhere along C, where $\Delta\sigma_{nn}$ and $\Delta\sigma_{nt}$ are the normal and tangential tractions to C determined from Problem (ii). This assumes that the interface crack is poised for continuing advance at every point along C and has nowhere become subcritical. This is a reasonable assumption for a decohering region spreading from a lengthening cut and is consistent with the shapes produced in Section 4. Alternatively, one could cut the film when $\sigma < \sigma_c$ and then increase σ above σ_c, for example by changing temperature and exploiting thermal expansion mismatch between the film and substrate. In this case, too, one can reasonably expect that the crack front will advance simultaneously everywhere along C as the temperature is changed. By (19), condition (17) can be rewritten as

$$\sigma_{nn}^2 + \frac{2\lambda}{(1-\nu)} \sigma_{nt}^2 = \sigma_c^2 \tag{20}$$

where σ_{nn} and σ_{nt} are the stresses in Problem (ii).

From dimensional considerations and the form of (20), it is readily concluded that the *shape* of the decohered region depends only on

$$\sigma/\sigma_c , \ \lambda \text{ and } \nu \tag{21}$$

while the *size* scales with the length of the cut, 2L.

3. LOCAL ANALYSIS AT THE END OF THE CUT

Observations of the cut test indicate that in some instances the decohesi' 'ndary C remains attached to the end of the cut while for some other material cumbinations   off the end of the cut. Insight into this issue is gained by investigating the conditions under which the

boundary can be attached. Consider the local geometry shown in Fig. 6 where the boundary emerges from the end of the cut at an angle ω. Let (r,θ) be planar-polar coordinates centered at the cut end, and look for a solution to the plane stress boundary value problem for the wedge region of Fig. 6 where the stresses depend on θ but not on r. Such a solution is readily found. Along the rigidly clamped edge on $\theta = \omega$ the stresses are found to be

$$\sigma_{\theta\theta} = -2\sigma \cos 2\omega [1 - \nu + (1+\nu)\cos 2\omega]^{-1}$$

$$\sigma_{r\theta} = -\sigma(1-\nu)\sin 2\omega [1 - \nu + (1+\nu)\cos 2\omega]^{-1} \tag{22}$$

$$\sigma_{rr} = \sigma[1 - \nu - (1+\nu)\cos 2\omega][1 - \nu + (1+\nu)\cos 2\omega]^{-1}$$

A separate calculation for the wedge region of Fig. 6 with homogeneous boundary conditions (i.e. with $\sigma_{\theta\theta} = -\sigma$ replaced by $\sigma_{\theta\theta} = 0$ on $\theta = 0$), shows that the stresses from the homogeneous problem will approach zero as $r \to 0$ as long as

$$1 - \nu + (1+\nu)\cos 2\omega > 0 \tag{23}$$

That is, the most singular admissible stresses to the homogeneous problem vanish as $r \to 0$ when (23) is met. (For $\nu = 1/3$, (23) requires $\omega < 60°$, while for $\nu = 1/2$, $\omega < 54.7°$.) Thus for a boundary emerging from the cut with an angle ω satisfying (23), the local stresses along the boundary on $\theta = \omega$ are given by (22).

Now consider the requirement that (20) be satisfied along $\theta = \omega$. With $\sigma_{nn} = \sigma_{\theta\theta}$ and $\sigma_{nt} = \sigma_{r\theta}$ from (22), (20) becomes

$$F(\omega,\lambda) = (\sigma_c/\sigma)^2 \tag{24}$$

where

$$F(\omega,\lambda) = [4\cos^2 2\omega + \lambda(1-\nu)\sin^2 2\omega]/[1 - \nu + (1+\nu)\cos 2\omega]^2 \tag{25}$$

A plot of F as a function of ω for various λ is shown in Fig. 7 for $\nu = 1/3$.

Next, recall that the regime of extensive decohesion under consideration requires $\sigma > \sigma_c$ so that the right hand side of (24) is less than unity. The angle ω made by the boundary satisfying (24) is given by the intersection of the solid-line curve in Fig. 7 (for a given λ) with the dashed-line curve for a given σ/σ_c, as illustrated in Fig. 7. The solution where the dashed-line

intersects the solid curve on the right where F is increasing with ω is unstable in the sense a slight increase in ω would perturb the front to a condition in which the left hand side of (20) would exceed σ_c^2. One notes immediately that there exists no solution when $\lambda = 1$, since $F \geq 0$. In other words, an attached boundary emerging from the end of the cut at some angle is not possible for an ideally brittle interface. By contrast, when $\lambda = 0$ there exists a solution for ω for all $\sigma > \sigma_c$. For $0 < \lambda < 1$, there exists an attached local solution for a range of σ/σ_c, but for σ/σ_c sufficiently large no such solution exists. The numerical results presented in the next section are completely consistent with the conclusions drawn from the local analysis. In fact, it wasn't until this local analysis was conducted that the numerical analysis could be formulated properly.

4. NUMERICAL RESULTS AND COMPARISON WITH SOME OBSERVATIONS

The numerical method used to solve for C made use of a sequence of interations wherein C was adjusted in a systematic manner until (20) was satisfied. In a given iteration with a given C, the plane stress boundary value problem (ii) of Fig. 4 was solved using a finite element procedure which subdivided the region within C into a grid of elements with due attention to achieving resolution near the ends of the cut. The normal and tangential tractions, σ_{nn} and σ_{nt}, on C were determined, the error in (20) was computed, and an adjustment to C for the next iteration was established. Details of this procedure are spelled out in Appendix II. Insight from the local analysis of the previous section was essential in knowing how to deal with the issue of whether C attaches or stands off the end of the cut. Under conditions where C is attached, the numerical prediction for the wedge angle ω was in excellent agreement with the prediction from the local analysis.

One-quarter sections of the computed decohered regions are shown in Fig. 8 for $\lambda = 0$, $1/2$ and 1, in each case for $\nu = 1/3$. When mode 3 has no influence on decohesion ($\lambda = 0$), the boundary of the decohered region always remains attached to the ends of the cut. When $\lambda = 1$, the boundary always stands off the end of the cut. Moreover, for a given value of σ/σ_c, the extent of the decohesion is more extensive when $\lambda = 1$ than when $\lambda = 0$. When $\lambda = 1/2$, the

local analysis predicts that the boundary is attached to the ends of the cut for $\sigma/\sigma_c < 1.1$ and stands off for $\sigma/\sigma_c > 1.1$, consistent with the boundaries shown in Fig. 8.

Decohered regions for a thin film of polyimide approximately 15μm thick and bonded to a glass plate approximately 2mm thick are shown in Fig. 9 and 10. Figure 9 shows decohesion spreading outside a circular cut of about .75cm in radius. (The thin film inside the cut has completely detached.) Figure 10 shows decohesion from a straight cut of approximately 1cm in length made on the same plate. Each of the two smaller decohered regions below the cut appear to be pinned at a point along the cut, probably because the cut was not as sharp as it should have been on that side. Note, however, that the three shapes are similar.

Information from the extent of the decohered region from the circular cut together with the observed shape of the decohered region in the straight cut test can be used to good advantage to draw conclusions about the nature of the decohesion criterion. For the axisymmetric decohesion boundary induced by the circular cut, $\sigma_{nt} = 0$, by symmetry, so that mode 3 plays no role. The plane stress problem for the circular cut is readily solved (Farris and Bauer (1989))[†]. Imposition of the criterion (20) then gives the extent of the decohesion. The ratio, R_0/R_i, of the radius of the decohered region to the radius of the circular cut is found to satisfy

$$\sigma/\sigma_c = \frac{1}{2}\left[(1-\nu)(R_0/R_i)^2 + 1 + \nu\right] \tag{26}$$

By measuring the ratio R_0/R_i in Fig. 9 and using $\nu = 1/3$, we determined $\sigma/\sigma_c \cong 1.21$ from (26). (Note that this procedure enabled us to avoid direct measurement of individual quantities such as $E, t, \mathcal{G}_c^*, \sigma$, etc.) Next, assuming $\sigma/\sigma_c = 1.21$ and $\nu = 1/3$, we computed the shapes of the decohered region for the straight cut as they depend on λ. These computed shapes are shown in Fig. 11. For $\sigma/\sigma_c = 1.21$ with $\nu = 1/3$, the transition from an attached boundary to one which stands off the ends of the cut occurs at $\lambda = .34$. The shape which best approximates the observed shape in Fig. 10 corresponds to $\lambda \cong .3$, nearly at the transition to

stand-off. With reference to the criterion (20), the value $\lambda \cong .3$ for the polyimide/glass interface suggests that mode 3 has a relatively small influence on the interface fracture process. Evidently the tip of the interface crack is heavily shielded from the full effect of mode 3. This finding is consistent with the analogous observations for an epoxy/glass interface under combined modes 1 and 2 (Cao and Evans (1988)). For that system, the tip of the plane strain interface crack was shielded from the effect of mode 2 by a comparable amount.

5. CONCLUDING REMARKS

The residual stress state assumed throughout this paper was equal biaxial tension σ. One can readily verify that the shapes predicted for the equal biaxial residual stress state ($\sigma_{11} = \sigma$, $\sigma_{22} = \sigma$) are unchanged according to the model if the residual stress component parallel to the cut σ_{11} is different from σ, i.e. for ($\sigma_{11} = \sigma_0$, $\sigma_{22} = \sigma$). Observations of decohesion under uniaxial stressing perpendicular to cut were recently made on a model system where a thin layer of transparent rubber adheres by van-der-Waals forces to a silicone rubber substrate (Ashby and Burwell, 1988).

The shape of the decohered region in the straight cut test is quite sensitive to the extent to which mode 3 enters the decohesion criterion. For a polyimide/glass combination the observed shapes suggest that the interface is far from being "ideally brittle" with significant shielding of the crack tip from the influence of mode 3. This study emphasizes the importance of mode-dependence in interface fracture. A family of decohesion criteria is proposed in Appendix I which is applicable to arbitrary combinations of modes 1, 2 and 3 and which can model the full range of behavior from ideally brittle to fracture controlled by K_1, as long as $K_1 > 0$. We suspect that the least certain aspect of the present model is the use of the criterion (17) when $\Delta\sigma_{nn} > 0$, as it is just opposite the ends of the cut when the decohesion boundary stands off the ends. As discussed in Section 2, the crack front will be closed when $\Delta\sigma_{nn} > 0$ with $K_1 = 0$ and it seems unlikely that (17) should continue to apply, at least not with the same values of λ and $\mathcal{G}_c^*$ as in the range when $\Delta\sigma_{nn} < 0$.

Further evidence of the complication which occurs when the film comes into contact with the substrate is seen in Figs. 13 and 14. The system in these two figures is also a polyimide film on a glass substrate but here the film is almost ten times thicker (t = 150μm) than in the system discussed in the previous section. The cut length is approximately 6mm. In these examples the film was cut at a temperature (T = 180°C) at which the residual stress initiates limited decohesion. The specimens are then cooled down. The thermal expansion mismatch between the film and the substrate increases the residual tension in the film as the temperature drops and decohesion spreads. The three decohesion regions shown in Fig. 13 are associated with the three approximate temperatures, T = 120°C, 80°C and 40°C, respectively. The decohesion boundary does stand off the ends of the cu' by a small amount and, moreover, the crack front is open ($K_1 > 0$) just opposite the cut ends. A second highly noticeable feature in Fig. 13 is the region of contact between the film and the substrate, as indicated by the dark spots more or less in the center of each of the decohered zones. These spots of contact appear to have had a significant influence on the shape of the decohered region, retarding its growth transverse to the cut. Note that the location and size of the contact spots changes as the temperature is lowered. On the upper side of the cut one can see the interaction and subsequent merging of the boundary of the decohered region with a small circular decohered region preexisting prior to the cut having been made.

There is evidence that a fairly substantial residual stress gradient exists in the film in this system with larger tensile stress at the interface than at the surface. A piece of fully decohered film curls consistent with the above observation. This residual stress gradient could account for the crack front remaining open at the ends of the cut and it might have some influence on the tendency of the film to contact the substrate.

Figure 14 shows another example for the same system cooled down to room temperature. The decohered region on one side of the cut is fully developed with no interior contact spots while that on the other side has been retarded by a substantial contact region. If contact occurs in the early stage of the process it seems to persist. Conversely, a decohering region free of interior

contact zones tends to avoid contact as it spreads. It is almost as if contact enters the process as an initial condition.

Acknowledgements

REFERENCES

Ashby, M. F. and Burwell, J. R. (1988). Preliminary results of delamination experiments. (private communication).

Cao, H. C. and Evans, A. G. (1989). An experimental study of the fracture resistance of bimaterial interfaces. To be published in Mechanics of Materials.

Evans, A. G., Drory, M. D. and Hu, M. S. (1988). The cracking and decohesion of thin films. *J. Mater. Res.* 3, 1043-1049.

Evans, A. G. and Hutchinson, J. W. (1989). Effects of non-planarity on the mixed mode fracture resistance of bimaterial interfaces. *Acta metall.* 37 (3), 909-916.

Farris, R. J. and Bauer, C. L. (1989). A self-delamination method of measuring the surface energy of adhesion of coatings. To be published in J. Adhesion.

Hutchinson, J. W. (1989). Mixed mode fracture mechanics of interfaces. To be published in Acta Met./Scripta Met. Proceedings of Conference on Metal/Ceramic Interfaces.

Kinloch, A. J. (1987). *Adhesion and Adhesives*, Chapman and Hall, London, p. 313.

Suo, Z. and Hutchinson, J. W. (1989). Interface crack between two elastic layers. To be published in Int. J. Fracture.

Suo, Z. and Hutchinson, J. W. (1989). Steady-state cracking in brittle substrates beneath adherent films. To be published in Int. J. Solids and Structures.

Thouless, M. D., Cao, H. C. and Mataga, P. A. (1989). Delamination from surface cracks in composite materials. J. Materials Science. (in press).

APPENDIX I

A FAMILY OF CRITERIA FOR INTERFACIAL FRACTURE UNDER MIXED MODE

Noting the expression (14) for the energy release rate of the interface crack under all three modes of loading (with $\beta = 0$), consider the family of interface toughness criteria specified by

$$\frac{1}{2}\left(\frac{1-v^2}{E} + \frac{1-v_s^2}{E_s}\right)(K_1^2 + \lambda_2 K_2^2) + \frac{\lambda_3}{4}\left(\frac{1}{\mu} + \frac{1}{\mu_s}\right)K_3^2 = \mathfrak{I}_c^0 \tag{27}$$

where $\mathfrak{I}_c^0$ can be regarded as the pure mode I toughness. The parameters λ_2 and λ_3 adjust the extent to which the shearing modes, K_2 and K_3, affect decohesion; they are limited to range between 0 and 1. It is assumed that $K_1 > 0$. For given values of λ_2 and λ_3, this criterion can be rewritten as

$$\mathfrak{I} = \mathfrak{I}_c(\psi, \phi) \tag{28}$$

where ψ and ϕ are Euler angles in the space (K_1, K_2, kK_3), as shown in Fig. 12, and where

$$\mathfrak{I}_c(\psi, \phi) = \mathfrak{I}_c^0[1 + (\lambda_2 - 1)\sin^2\psi \, \sin^2\phi + (\lambda_3 - 1)\cos^2\phi]^{-1} \tag{29}$$

In casting the criterion in the form (28), it has been mathematically convenient to scale the K_3-axis by k where

$$k^2 = \frac{1}{2}\left(\frac{1}{\mu} + \frac{1}{\mu_s}\right)\left[\frac{1-v^2}{E} + \frac{1-v_s^2}{E_s}\right]^{-1} \tag{30}$$

This family of criteria includes the following special cases.

(i) "Ideally brittle" decohesion:

$$\lambda_1 = \lambda_2 = 1 \quad , \quad \mathfrak{I}_c = \mathfrak{I}_c^0 \tag{31}$$

(ii) Decohesion independent of the shearing modes (i.e. $K_1 = K_1^c$):

$$\lambda_1 = \lambda_2 = 0 \quad , \quad \mathfrak{I}_c = \mathfrak{I}_c^0[1 - \sin^2\psi \, \sin^2\phi - \cos^2\phi]^{-1} \tag{32}$$

where $\mathcal{S}_c^0$ is the pure mode 1 toughness related to K_1^c by (27) with $K_2 = K_3 = 0$.

(iii)   The decohesion criterion (17) used in the body of the present paper is included in (28) if one makes the identification

$$\lambda = \lambda_3[1 + (\lambda_2-1)\sin^2\psi]^{-1} \quad , \quad \mathcal{S}_c^* = \mathcal{S}_c^0[1 + (\lambda_2-1)\sin^2\psi]^{-1} \tag{33}$$

Thus, a solution in the body of the paper for a given pair $(\lambda, \mathcal{S}_c^*)$ corresponds to a solution for any combination of $(\lambda_2, \lambda_3$ and $\mathcal{S}_c^0)$ satisfying (33). For example, the solution for $\lambda = 0$ applies to an interface with $\lambda_2 = \lambda_3 = 0$ and $\mathcal{S}_c^* = \mathcal{S}_c^0$, but it also applies to any interface with $\lambda_3 = 0$ and $\lambda_2 \neq 0$ if $\mathcal{S}_c^* = \mathcal{S}_c^0[1 + (\lambda_2-1)\sin^2\psi]^{-1}$.

A comparison of (27) with the Cao-Evans (1989) mode 1/mode 2 data for epoxy on glass is shown in Fig. 15. The choice $\lambda_2 = .15$ gives a reasonable fit to this data. Now suppose that the value $\lambda_2 = .15$ applies as well to the polyimide/glass system used in the present study. With $\lambda = .3$ inferred from the shape of the decohesion in the cut test and with $\psi = 50°$, Eq. (33) gives $\lambda_3 = .15$. Thus it would appear that the polyimide/glass interface crack is heavily shielded from both mode 2 and mode 3 by comparable amounts.

The criterion (27) can be expressed in a form which has been considered in the adhesives literature (Kinloch, 1987). When $\beta = 0$, separate "components" of the energy release rate can be meaningfully defined as

$$(\mathcal{S}_1, \mathcal{S}_2) = \frac{1}{2}\left(\frac{1-\nu^2}{E} + \frac{1-\nu_s^2}{E_s}\right)(K_1^2, K_2^2)$$

$$\mathcal{S}_3 = \frac{1}{4}\left(\frac{1}{\mu} + \frac{1}{\mu_s}\right)K_3^2 \tag{34}$$

Then it is immediately noted that (27) can be rewritten as

$$\frac{\mathcal{S}_1}{\mathcal{S}_1^c} + \frac{\mathcal{S}_2}{\mathcal{S}_2^c} + \frac{\mathcal{S}_3}{\mathcal{S}_3^c} = 1 \tag{35}$$

where $\mathcal{S}_1^c = \mathcal{S}_c^0$, $\mathcal{S}_2^c = \mathcal{S}_c^0/\lambda_2$ and $\mathcal{S}_3^c = \mathcal{S}_c^0/\lambda_3$. In this form, $\mathcal{S}_2^c$ and $\mathcal{S}_3^c$ are regarded as the pure mode toughnesses. The intention behind (27) is that mode 2 and mode 3 act together with

mode 1. It is not necessarily expected that (27) should accurately span the whole range of behavior to include pure mode 2 or mode 3.

APPENDIX II

NUMERICAL METHOD

The plane stress problem is Problem (ii) of Fig. 4 with boundary conditions (12). The boundary C of the decohered region must be found such that (20) holds along C where σ_{cn} and σ_{ct} are the normal and tangential stresses to C from Problem (ii).

For a given shape, the stresses are determined using the finite element method. A coarse mesh (4×3) is shown in Fig. 16a. Typically, a (30×30) mesh gave results of sufficiently high accuracy. Each quadrilateral consists of two linear displacement elements with straight edges.

Two methods for representing a boundary shape that locally satisfies the decohesion criterion have been compared. Both methods lead to an iterative procedure.

In the first method, the boundary curve C is represented by a set of functions according to

$$y(x) = a_0 + a_1 f_1(x) + a_2 f_2(x) + ... + a_N f_N(x) \tag{36}$$

Let $x = 0$ be at the end of the cut and $x = L$ at the symmetry line at the center of the cut as indicated in Fig. 16a. The following polynomial expansion is used

$$f_1(t) = t(2-t) \ , \ f_2(t) = t^2(3-2t) \ ,$$

$$f_n(t) = t^2(t-1)^2 T_{n-3}(2t-1) \qquad 3 \leq n \leq N \tag{37}$$

where $t = x/L$ and T_n are Chebyshev polynomials of order n.

For the case where the decohesion zone stays attached, $a_0 = 0$ in (36) and the results of the FEM analysis and the local analysis of Section 3 can be compared. The angle between the cut and the decohesion zone as given by the two methods is usually within 0.5%.

In the second method, the boundary curve is represented by a piecewise linear curve as in the FEM approximation in Fig. 16a. The boundary is in this way described by a set of coordinates (x_p, y_p), $p = 1, 2, ..., P$, with a linear variation between these points.

Let the initial boundary curve be represented by polynomials as in the first method above. The criterion (20) is approximately satisfied along C by choosing the coefficients a_i to minimize

$$E_{\pi} = \int_C \left[\sigma_c^2 - \left\{ \sigma_{ba}^2 + \frac{2\lambda}{(1-\nu)} \sigma_{at}^2 \right\} \right]^2 ds \tag{38}$$

The derivatives and curvatures of E_{π} with respect to each of the a_i are obtained numerically by solving the FEM problem for $a_i + \delta a_i$ and $a_i - \delta a_i$. In an iteration, new values of a_i are found by choosing δa_i to minimize E_{π} according to the derivatives and curvatures so obtained. In this way, the iteration process usually converges in 5 iterations.

Now let the boundary curve C be represented by the finite element grid as in the second method above. The stress change for a small movement of a nodal point is found for each node on the boundary curve. Use is made of the fact that the FEM grid is unchanged except in three elements, as shown in Fig. 16b, resulting in a nearly unchanged stiffness matrix. The nodes at the boundary are then moved to satisfy the decohesion criterion (20) in each element at the boundary. Since the stress change at C is nonlinear with respect to movements of the nodes of C, the process is iterative. Nodes are moved with x_p kept fixed except when y is small close to the ends of the cut where y_p is kept fixed.

The two methods outlined above result in essentially identical shapes of the decohesion zone. Typically, $N = 4$ in (36) is sufficient, since results for $N = 6$ usually are indistinguishable. In the cases when the decohesion zone stands off the end of the cut, the set of functions (37) is used in (36) for $0 \le x \le L$. For $x < 0$, C is represented by

$$y(x) = c_1 (x - c_2)^{1/2} \tag{39}$$

with the c_i chosen to ensure continuity and smoothness of C at $x = 0$. The symmetry condtion at $y = 0$ is then fulfilled. The procedure employed first involved minimizing E_{π} with respect to the a_i's. Then the nodal points near the end of the cut are adjusted according to the second method in order to lower the error on this part of the curve.

No special treatment of the singular behavior at the ends of the cuts was made, other than focussing the FEM grid towards the end of the cut and comparing results for different levels of discretization.
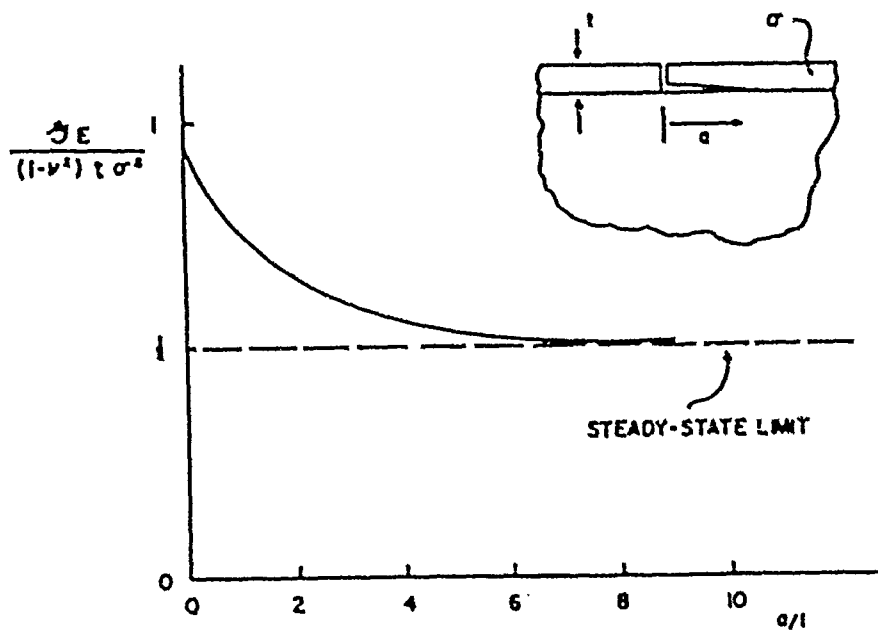
Fig. 1 Sketch of the variation of energy release rate $\mathcal{G}$ for a plane strain interface crack spreading from a cut in a film with residual biaxial tension σ. See Thouless, *et al.* (1989) for an accurate plot.
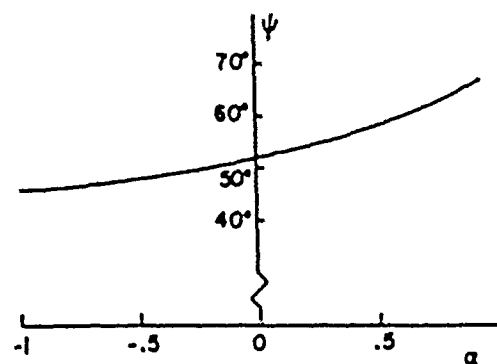

Fig. 2 Dependence of the phase of the stress intensity factors, $\psi = \tan^{-1}(K_2/K_1)$ for a steady state plane strain interface crack as a function of the elastic mismatch parameter α when $\beta = 0$. The crack is driven by residual tensile stress in the thin film.
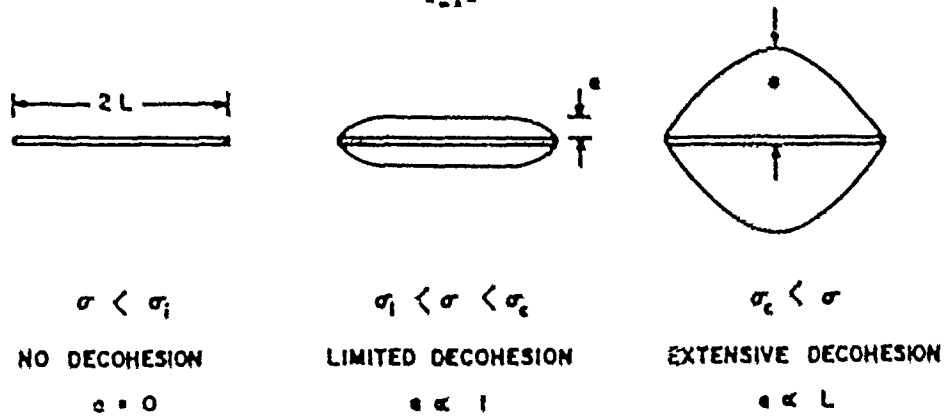
Fig. 3 Three regimes of behavior for cut test when $\sigma_i < \sigma_c$.
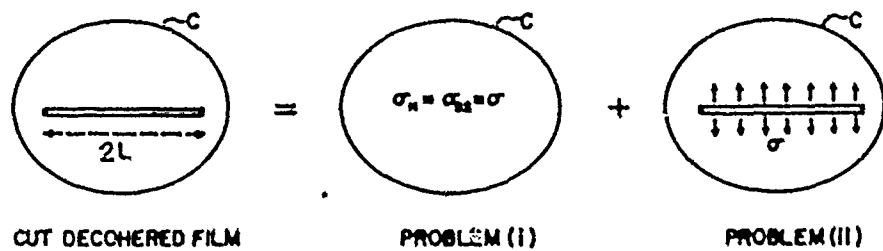
Fig. 4 Plane stress model of extensive decohesion when $\sigma > \sigma_c$. Problem (i) is state of uniform biaxial tension σ. Problem (ii) has clamped boundary C with normal traction σ applied to cut.
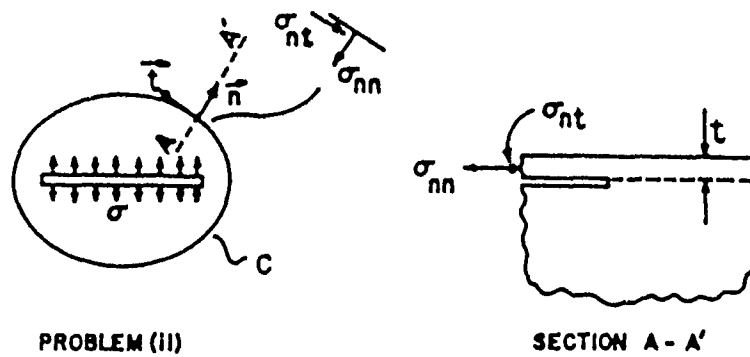
Fig. 5 Local tractions σ_{nn} and σ_{nt} in film along interface crack front for Problem (ii).
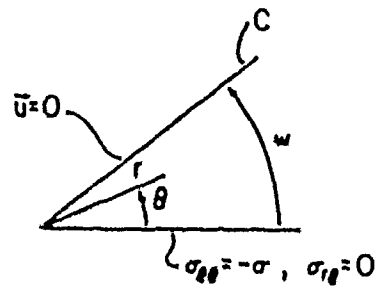
Fig. 6 Geometry at end of cut for local analysis of attached boundary.
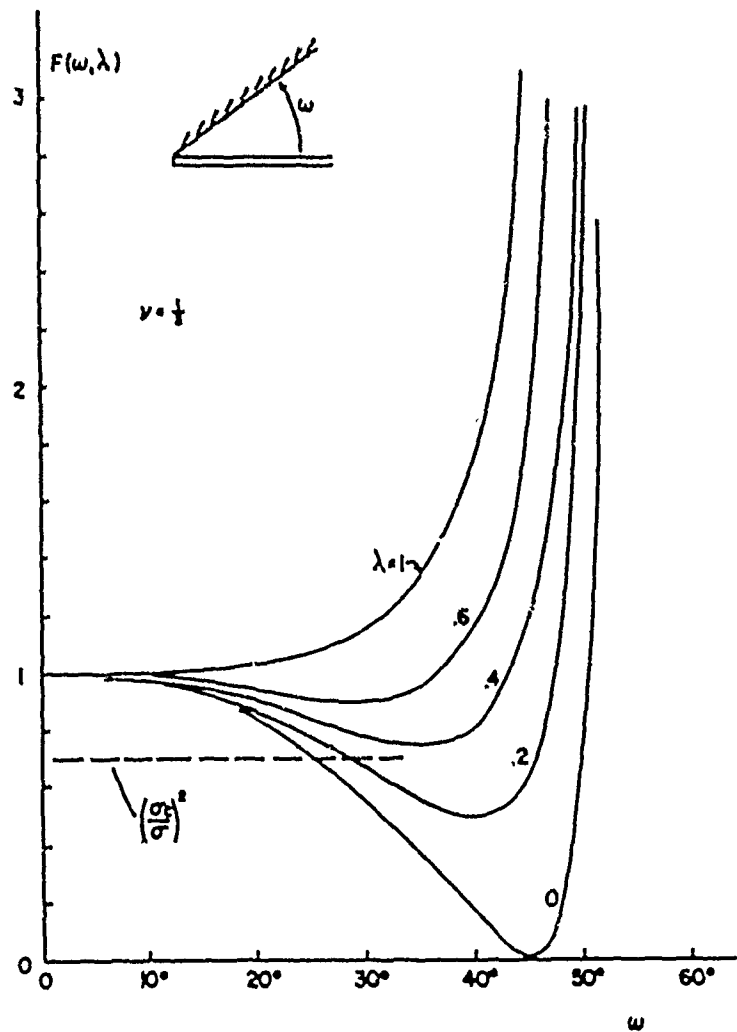
Fig. 7 Function $F(\omega,\lambda)$ associated with local analysis of boundary attached to end of cut. Condition for determining ω is $F = (\sigma_c/\sigma)^2$
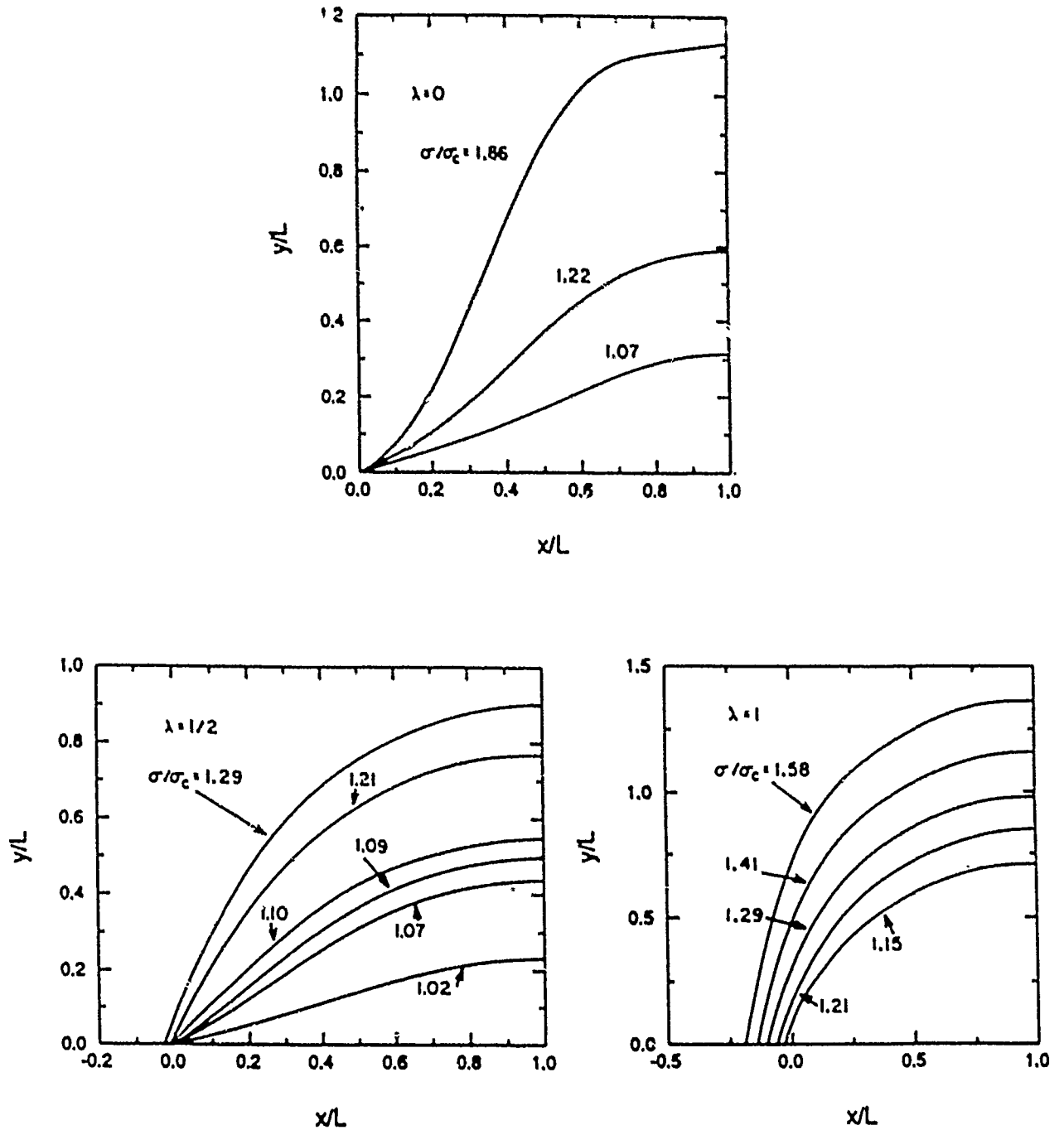
Fig. 8 Computed boundaries of decohered region for various values of σ/σ_c for $\lambda = 0$, $1/2$ and 1 with $v = 1/3$. The case $\lambda = 0$ applies to a criterion based on $K_1 = K_1^c$ along the boundary independent of K_2 and K_3, while the case $\lambda = 1$ applies to an ideally brittle interface where $\vartheta = \vartheta_c^0$ along the boundary.

Fig. 9 Decohered region exterior to a circular cut on a glass plate with a film of polyimide.
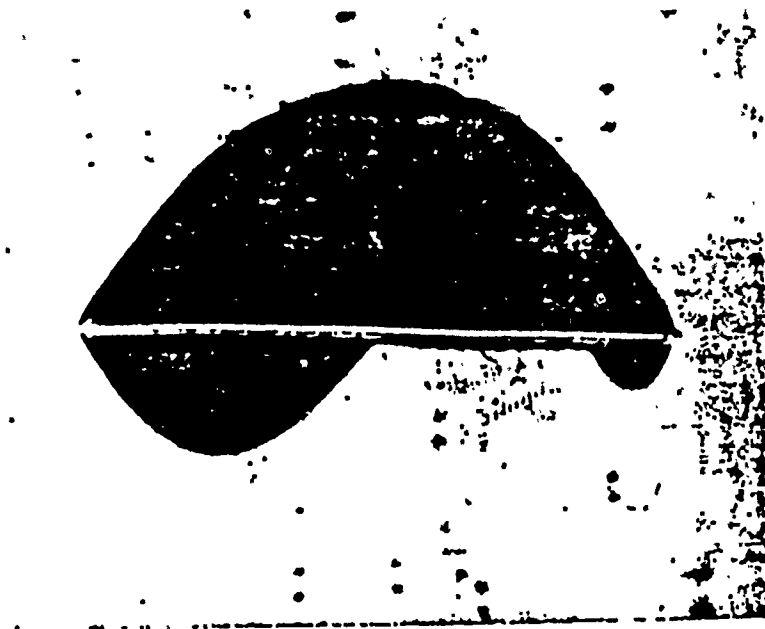
Fig. 10 Decohered regions from a straight cut on a glass plate with a film of polyimide. The plate is the same as that for the circular cut in Fig. 9.
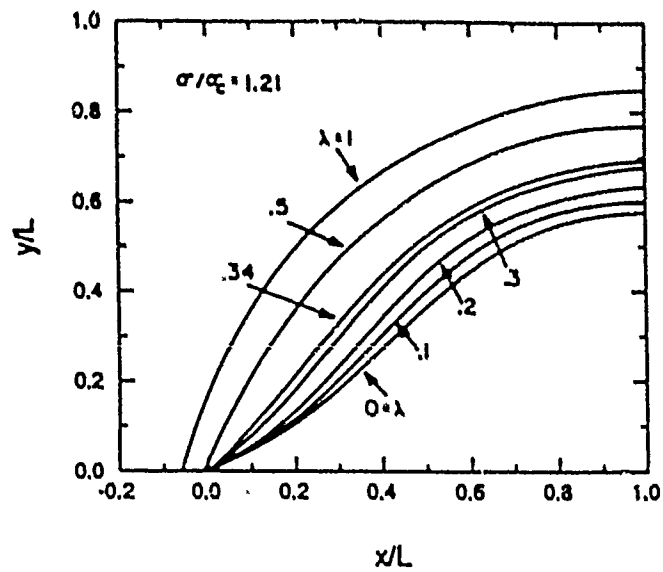
Fig. 11 Computed boundaries of decohered boundaries for $\sigma/\sigma_c = 1.21$ and $v = 1/3$ for various values of λ_0. The value $\sigma/\sigma_c = 1.21$ corresponds to the value of the parameters for the polyimide film/glass plate system in Figs. 9 and 10.
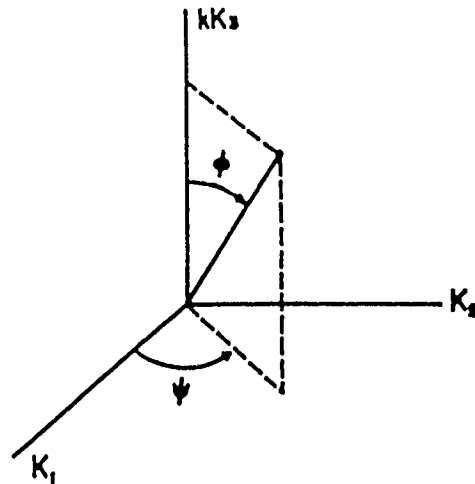
Fig. 12 Definition of Euler angles in (K_1, K_2, kK_3)-space.
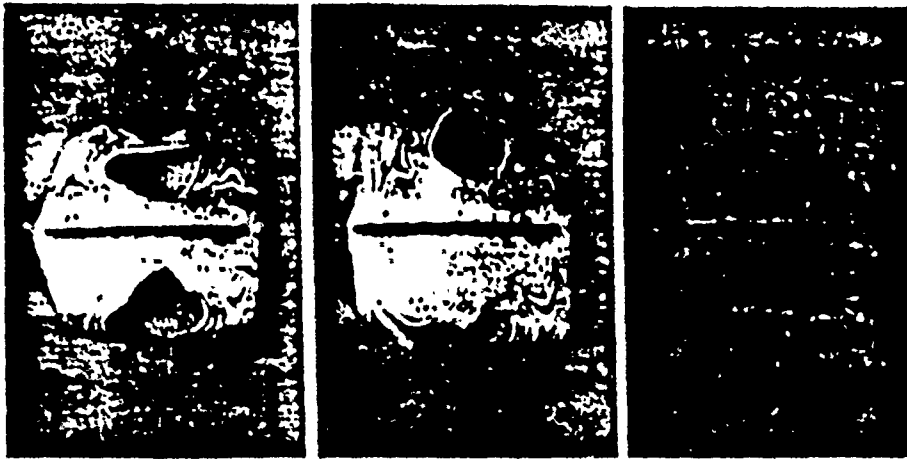
Fig. 13 Decohered regions from cut for a relatively thick polyimide film on glass at three temperatures: T = 120°C, 80°C and 40°C. Regions where the film has rebonded to the substrate are evident.

Fig. 14 Decohered region for the relatively thick polyimide film on glass at room temperature.
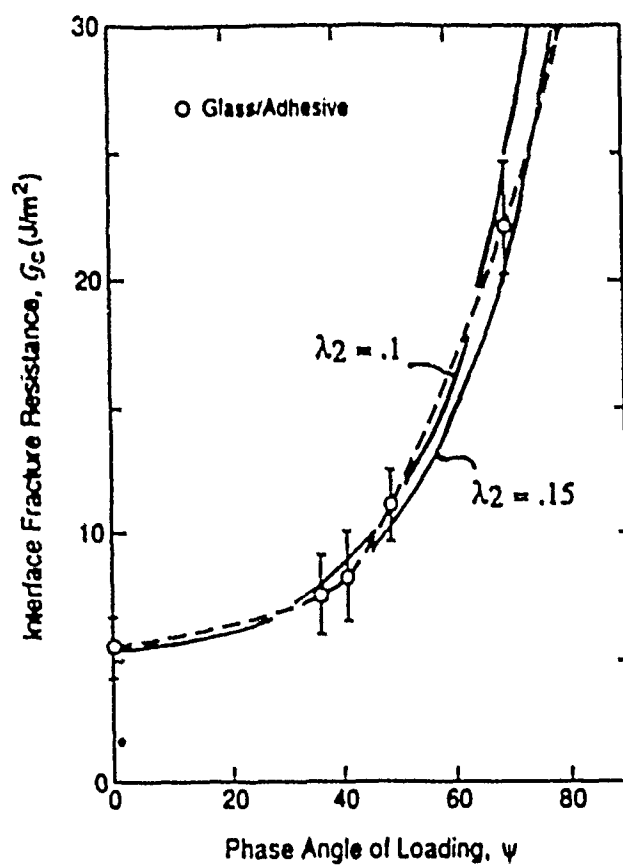
Fig. 15 Comparison of the Cao-Evans (1989) data for an epoxy/glass interface with the phenomenological criterion (27) for two values of λ_2.
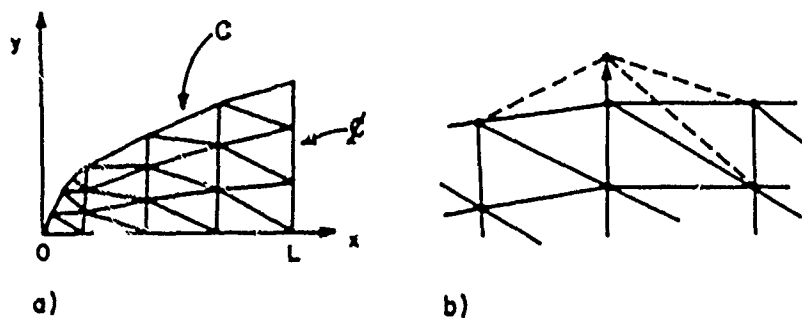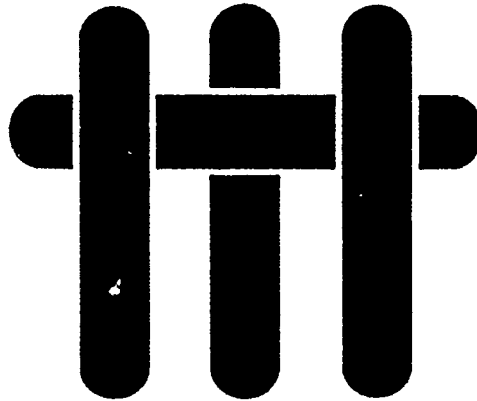

Fig. 16 (a) Finite element grid. (b) Movement of boundary node.

MATERIALS

MEASURING THE STRENGTH AND STIFFNESS OF THIN FILM MATERIALS BY MECHANICALLY DEFLECTING CANTILEVER MICROBEAMS

T. P. Weihs, S. Hong, J. C. Bravman and W. D. Nix

Department of Materials Science and Engineering
Stanford University
Stanford, California 94305

MEASURING THE STRENGTH AND STIFFNESS OF THIN FILM MATERIALS BY MECHANICALLY DEFLECTING CANTILEVER MICROBEAMS

T. P. WEIHS, S. HONG, J. C. BRAVMAN and W. D. NIX

Department of Materials Science and Engineering
Stanford University, Stanford, CA, 94305.

Abstract

The present authors recently introduced the mechanical deflection of cantilever microbeams as an experimental technique for measuring the strength and stiffness of thin films.[1] The technique utilizes conventional integrated circuit (IC) fabrication to process the samples, a Nanoindenter to deflect the cantilever beams mechanically, and simple elastic beam theory to analyze the data. This paper will review the technique and describe some of its current applications.

Introduction

When measuring the mechanical properties of thin films, the presence of substrates below the films can affect the measurements[2]. Thus, one would prefer to test the films free of their substrates. However, if the films are approximately 1.0 μm thick and free-standing, they can be damaged if handled directly. To avoid both of these problems, and to offer a measurement of the strength and stiffness of thin films on a local scale, the deflection of cantilever microbeams using a Nanoindenter was introduced recently[1]. This paper will describe the three basic stages of the microbeam technique - the fabrication of samples, the testing of samples, and the analysis of data. In addition, some applications of the technique will be described and several issues regarding the deformation of the beams will be discussed.

Sample Fabrication

A variety of different materials can be fabricated into cantilever beams for testing. Interconnect metals such as gold, dielectric materials like silicon dioxide, and semiconductors such as silicon have all been tested. For such a diverse group of materials, different IC processing methods are employed.

The first method -- called the "lift-off" technique -- is typically used in the processing of metal beams. It is termed "lift-off" because photoresist is peeled away or "lifted off" the silicon substrate in order to pattern the metal film. The initial step in the technique is the deposition and patterning of photoresist. A positive resist is spun onto a lightly-doped p-type (100) silicon wafer and patterned using typical exposures to UV light and chemicals. The processing leaves photoresist only in sections where the metal film is not desired. In the next step, metal is deposited over the complete top side of the wafer using e-beam evaporation or sputtering techniques. After the deposition, warm acetone dissolves the photoresist that is covered by the metal film and thereby lifts off the unwanted sections of the film. The cantilever beams then appear as thin lines on the silicon. To form the bottom side of the cantilever beams, a heated solution of EDP (66 wt.% ethylendiamine, 23 wt.% H_2O, 11 wt.% pyrocatechol) anisotropically etches away the silicon from below the metal lines. The final sample will appear as shown in Figure 1: cantilever beams extending over an open silicon etch pit that is bounded by 4 {111} side walls and a (100) bottom surface.

The second fabrication method, used for fabricating dielectric samples, is more common in IC processing. Silicon dioxide films are grown (wet-thermal oxide) or deposited (low-pressure chemical vapor deposition) on (100) Si wafers. The oxide films are then patterned using standard photolithography processes that include positive photoresist and a wet etchant of buffered hydrofluoric acid. Once the oxide is shaped into thin lines, a heated solution of EDP etches away the Si from beneath the patterned oxide films, thereby forming cantilever beams of SIO_2.

The third and final processing technique is used to fabricate silicon beams. It is significantly different than the first two in that we etch the silicon wafer from its back side to form the cantilever beams and not from its front or top side. To begin the processing, silicon nitride is deposited on both sides of the wafer. 3.0 mm squares are then patterned in the nitride on the back side of the wafer using photoresist and dry etching. After the photoresist is removed, liquid KOH anisotropically etches the silicon wafer until the square patterns extend up through most of the wafer. By carefully monitoring the depth of the square etch pit, the etching can be stopped so that only a thin film remains at the top of the wafer. In the final set of steps, these thin films of silicon are patterned into cantilever beams. First, the top of the wafer is patterned with a hardened photoresist. Then, sections of the nitride film are dry etched where silicon is to be removed. Once the silicon is etched to form the beams, the remaining nitride and photoresist layers are removed as well.
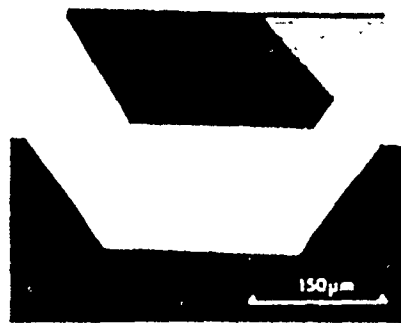

Figure 1: Au Cantilever Microbeams over a Si Etch Pit

Mechanical Testing

The elastic and plastic properties of the various materials comprising the thin cantilever beams are measured using a Nanoindenter. The instrument mechanically deflects the beams at a constant velocity while continuously recording the applied load and the measured deflection. The rates of loading and unloading are controlled so that the beams bend and recover at a constant velocity between 3 and 6 nm/s. During loading, when a predetermined deflection is reached, the applied force is held constant for 10 s before it is removed. The resolutions of load and displacement measurements are 0.25 µN and 0.3 nm, respectively.

To analyze the load and deflection data recorded for a given test, the physical dimensions of the beam must be known. The width and thickness of a beam are typically measured using scanning electron microscopy and optical microscopy. However, to calculate mechanical properties we do not measure the actual length of the beam. Instead, we measure the distance between the fixed end of the beam and the point of loading which we call the effective length. In Figure 2 the distance between point B and C marks this dimension. For a typical test, it is measured using the X and Y stepper motors on the Nanoindenter which have a 0.5 µm lateral resolution. These motors control the movement of the sample and they prerecord positions A, B, and C. Of course, it would be more accurate to measure the effective length directly by looking for an indentation on the beam at the point of contact. However, the forces applied to the beams are typically too small to form a visible indentation on the top of the beam. As an alternative, we check the position of contact on the beam indirectly by observing the position of the indentation at point A. If the indentation is not at its prerecorded position, then changes in the point of loading on the beam can be inferred.

Figure 2 also shows a schematic of the main electro-mechanical components of the Nanoindenter. A brief description of these components is given here to clarify the test technique

and the presentation of data. In simplest terms, the Nanoindenter is a constant-load machine: loads are applied and displacements are measured. We can see from Figure 2 that load is applied to the shaft of the indenter by changing the DC current running through the coil. Changes in the current alter the magnetic field emanating from the coil, and thereby increase or decrease its attraction to the permanent magnet. Note that as the indenter shaft moves up or down, the applied load must offset the resistance of the suspending spring as well as the resistance of the beam being bent. Thus, once the indenter tip contacts a beam and begins to bend it, the applied load is working to deflect two springs in parallel: the suspending spring and the cantilever beam. A final point to notice regarding the electro-mechanical workings of the Nanoindenter is the effect of the relative magnitude of these two springs. If a beam is very compliant because it is thin or long, then its spring constant may be too small, relative to the spring constant of the suspending spring, to be measured accurately.

By continuously recording the load applied to the shaft of the indenter and the resulting deflection of the shaft, the initial contact between the diamond tip and the cantilever beam can be detected. In Figure 3 the total applied load is plotted versus the total tip deflection for two different tests of a silicon beam. In the initial parts of both curves, the slope or spring constants are equivalent, since no beam has yet been contacted. The applied load is deflecting only the suspending spring. At a total deflection of approximately 2.0 μm, the slopes of the two curves change abruptly because the indenter contacts the beams. The plot of data following the contact point then describes the total applied load and the total deflection of the indenter tip. To know the load on the beam, the force carried by the suspending spring must be subtracted, and to know the deflection of the beam, we must subtract the depth to which the diamond tip indents the top of the beam.
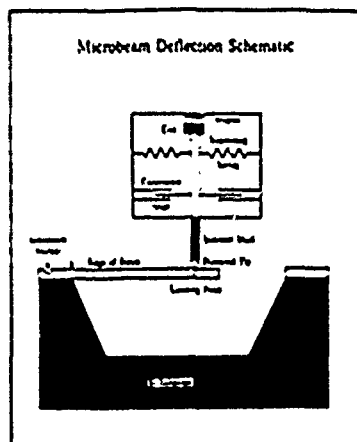

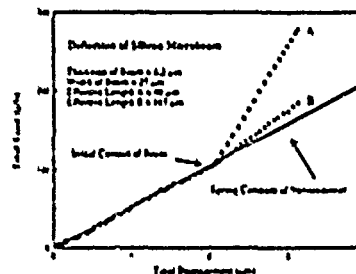Figure 2: Nanoindenter and Microbeam Schematic   Figure 3: Two Deflections of a Si Beam

Analysis of Beam Deflections

The total load applied by the instrument deflects both the suspending spring and the beam. To subtract the load on the suspending spring from the total load we simply measure the spring constant and include it in the analysis. To determine the net deflection of the beam, though, more work is required. For the softer, metallic beams, measurable indentations can form on the top of the cantilever beams as the forces are applied. These indentations are often too small to see, but they can be detected given the displacement resolution of the Nanoindenter. For

the harder, dielectric materials, the indentations are typically very small and appear to be completely elastic. This is suggested by data presented earlier[1] that shows complete recovery on unloading. However, whether the indentations are small or large, plastic or elastic, their depth should be subtracted from the total deflection data to analyze the deformation of the beams.

To subtract the depth of indentation from the total deflection data, an empirical equation relating the depth of indentation to the applied load is determined for each new beam material. An array of small indentations is made in sections of the material that are supported by the silicon substrate. The indentations vary in depth from 20 nm to 500 nm. To determine the empirical equations, curves are fit to the average values for indentations of similar depths.

Another deflection which must be considered is that of the cantilever beams bending across their width. Since the beams are loaded at a point along their centerline, and since the beams are typically very wide compared to their thicknesses, the centers of the beams will tend to bow downward relative to the edges of the beams. The authors have shown that for an effective length to width ratio of 1.0, the additional deflection of the diamond tip due to curvature across the widths of the beams can be 2.0 % of a beam's net deflection [1]. However, as the length or thickness of a beam increases compared to its width, this deviation decreases significantly. To improve the accuracy of the measurements, this displacement is considered in all analyses.

Once the net deflections and loads for the bending of a beam are known, simple elastic beam theory for plane strain is applied to calculate elastic and plastic properties. Elastic theory says that the applied load, P, should be linearly related to deflection, w, and that the slope of this relationship is determined by the Young's Modulus, E, of the beam material and the geometry of the beam.

$$w = \frac{4Pc^3}{bEt^3}(1 - v^2) \tag{1}$$

In the above equation c is the effective length, b is the width, t is the thickness, and v is the Poisson's ratio for the material. The two different deflections of a Si beam which are plotted in Figure 3 support equation 1. The Young's Modulus of the Si can be determined from the slopes of the two curves and their respective geometries. Note that the beam test with the longer effective length has the lower slope, as expected from equation 1. Furthermore, note that the linear nature of the curves in Figure 3 suggest that the two deflections are purely elastic.

In order to measure plastic properties such as yield strength, fracture strength, or strain hardening, beams can be deflected beyond their elastic limit. The onset of yielding or fracture is easily detected through a deviation from linearity in the load-deflection curves. The yield stress, σ_y, is given by the following equation

$$\sigma_y = \frac{6cP_y}{bt^2} \tag{2}$$

where P_y marks the load corresponding to the loss of linearity and the onset of plastic deformation. Note that equation 2 assumes the existence of plane stress rather than plane strain. This assumption is made for two reasons. The first is that the plane strain conditions assumed in the elastic analysis for equation 1 are relaxed at the edges of a beam. The second is that because of this relaxation, yielding should start at the edges of the beam where the effective stresses[3] are highest and then move inwards toward the center of the beam.

Discussion

To date, various metal beams have yielded and shown strain hardening (Figure 4). However, no beam has fractured at its fixed end even though strains in excess of 0.01 have been applied to silicon dioxide and silicon beams (Figure 5). The absence of such fracturing has been attributed to the uniformity of the material. Large planer defects, such as cracks, appear to be absent. The nonlinearities of the curves in Figure 5 do not mark yield or fracture. Instead they are attributed to the indenter tip sliding out along the hard oxide beam as it is bent. Such sliding

increases the effective length of the beam and thereby decreases the total spring constant for the test. Consequently, the slopes of the curves decrease. The deformation is thought to be elastic since the unloading curve retraces the loading curve.

When yielding does occur in a cantilever beam, the presence of surface layers such as oxides must be considered. As a cantilever beam is bent beyond its elastic limit, the material at the fixed end of the beam will be the first to deform plastically. The deformation will start at the top and the bottom surfaces of the beam, and at its edges, and then it will spread inwards. If the beam is completely homogeneous, the onset of plastic deformation at the surfaces of the beam will match the strength of the material within the beam. However, if a surface layer such as an oxide is present, then the initial yielding may be delayed and the measured yield strength may not accurately describe the beam material. This concern has been raised as we test metal beams such as Cu and Al which readily form oxide layers when exposed to air at room temperature.
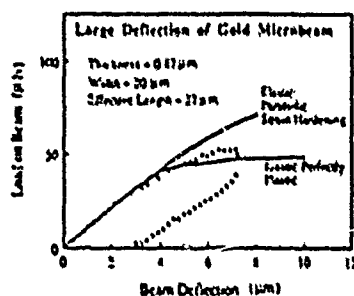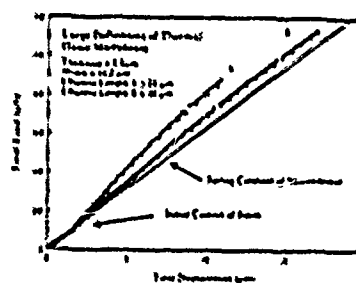

Figure 4: Large Deflection of a Au Beam

Figure 5: Large Deflection of a SiO₂ Beam

To estimate the degree of strain hardening that occurs in a metal beam as it is bent, we compare experimental data with theoretical curves. In Figure 4 the strain hardening of a Au beam is plotted with theoretical data for an elastic-plastic material (i.e. no strain hardening) and a material that strain hardens parabolically [1]. Since the experimental curve for Au lies closest to the curve for the elastic-plastic material, we assume that the thin film does not strain harden significantly during bending. The simple explanation for this is the large ratio of surface area to volume for the Au material. With the large area of free surface, dislocations can leave the beam and avoid strain hardening interactions. For a typical bulk sample that has a lower ratio of surface area to volume, dislocations cannot leave the material as easily. Thus, the dislocations interact more than in the Au beam, and strain hardening is higher.

For comparison of measurement techniques, the hardness and Young's Modulus of the films that form the beams were measured using a Berkovich Indenter, the Nanoindenter, and the indentation analysis of Doerner and Nix[2]. Arrays of indentations were made in sections of the films that are supported by the silicon substrates. Properties measured by both the microbeam deflection technique and the nanoindentation technique are presented in Table 1. The values of Young's Modulus measured by the two methods for amorphous and nontextured polycrystalline materials (SiO₂ and Au) should be directly comparable. However, in the initial experiments reported here, there was only limited agreement. The moduli measured by indentation for the oxide films are higher than those measured by microbeam deflection and they are higher than literature values reported earlier[1]. Some of this disparity can be attributed to a densification of the oxide films that appears to have occurred upon indentation[1]. In terms of the Au films, preliminary X-ray studies show little texturing in the Au and therefore suggest that the two measurement techniques should yield similar values for Young's Modulus. However, the

Indentation values are again higher. A possible cause of this difference and the difference in the oxide measurements is the effect of the large pressure under the indenter tip[1,2]. The only strong agreement between the measurements of the two techniques is shown for the plastic properties of the Au films. With an appropriate conversion from yield strength to hardness, $H = 3.2 \cdot \sigma_y$ [7], the two different measures of the resistance to plastic deformation compare very favorably.

When single crystal films such as silicon are tested, the microbeam technique measures Young's Modulus for a particular direction in a particular plane, while the indentation technique measures an average Young's Modulus[4-6]. Thus, the two techniques can not be compared directly. But, the Young's Modulus of 163 GPa measured for Si on a (100) plane and in a [110] direction is very similar to the calculated value of 169 GPa[8].

Table I. Elastic and Plastic Properties Measured for Thin Films

	Beam Test		Indentation Test		
	E	σ_y	E	H	v
Si	163	>3.39	155	13	.068(.20)²
Thermal SiO2	64	>0.57	85	12	.16
LTO	44	•	53	10	.16
Au	57	0.341	74	1.0	.42

*All data in GPa
1Corrected from initial reported value[1] in which an effective stress was mistakenly listed.
2An average value of v (0.20) was used for the indentation analysis.

Conclusion

The deflection of cantilever microbeams has been described as an alternative technique for measuring mechanical properties of thin film material. The various steps involved in sample fabrication, beam deflection, and data analysis have been reviewed and some current applications of the technique have been discussed.

Acknowledgements

References

1. T.P. Weihs, S. Hong, J.C. Bravman, and W.D. Nix, J. Mater. Res. 3, 931 (1988).
2. M.F. Doerner and W.D. Nix, J. Mater. Res. 1, 601 (1986).
3. G.E. Dieter, Mechanical Metallurgy, (McGraw-Hill, San Francisco, 1986), pp. 77-81.
4. C.J. Farthing, T.P Weihs, D.W Kisker, J.J. Krajewski, M.F. Tang, and D.A. Stevenson, presented at Fall MRS Meeting, Boston, MA, 1988 (to be published).
5. W.C. Oliver, F.A. List, and R.A. McKee, presented at the Fall MRS Meeting, Boston, MA, 1988 (to be published).
6. J.B. Pethica, presented at the Fall MRS Meeting, Boston, MA, 1988 (to be published).
7. A Bhattacharya and W.D. Nix, Int. J. Solids Structures, 24, 881 (1988).
8. W.A. Brantley, J. Appl. Phys., 44, 534 (1973).

THE DETERMINATION OF MECHANICAL

PARAMETERS AND RESIDUAL STRESSES

FOR THIN FILMS USING

MICRO-CANTILEVER BEAMS

S. Hong, T. P. Weihs, J. C. Bravman and W. D. Nix

Department of Materials Science and Engineering
Stanford University
Stanford, California 94305

THE DETERMINATION OF MECHANICAL PARAMETERS AND RESIDUAL STRESSES FOR THIN FILMS USING MICRO-CANTILEVER BEAMS

S. HONG, T. P. WEIHS, J. C. BRAVMAN AND W. D. NIX
Dept. of Materials Science and Engineering, Stanford University, Stanford, CA. 94305

ABSTRACT

A method for determining mechanical parameters and residual stresses for thin films is described. Multi-layer cantilever beams (LPCVD SiN$_x$/thermal SiO$_2$) are fabricated utilizing standard IC processing technologies and micromachining of silicon. The elastic response of the beams to imposed deflections is then measured using a Nanoindenter, a sub-micron hardness testing machine. The elastic constants of the nitride films are calculated from the force vs deflection slope and known elastic constants of the thermal SiO$_2$ and silicon. By measuring the curvature of the multi-layer cantilever beams with a scanning electron microscope after successive etching of the LPCVD nitride films, average and differential stresses in the films were calculated.

INTRODUCTION

LPCVD (low pressure chemical vapor deposition) SiN$_x$ films are widely used in silicon semiconductor devices as a diffusion barrier and as passivation layers. The structure of thin LPCVD SiN$_x$ over thermally grown SiO$_2$ is particularly important in the fabrication of most MOS devices, since this structure is essential for the LOCOS process. The residual stress in the nitride film is one of the main factors responsible for the generation of point and line defects in the silicon substrate during/after the bird's beak formation. The ratio of the thicknesses in the nitride and oxide layers is critical to the defect formation [1]. To gain a better understanding of the mechanism of defect generation, the stress state and elastic constant of the thin nitride films should be known.

In this paper, a new test method which can determine both the elastic constants and residual stresses of very thin films will be described. The technique is an extension of the single-layer cantilever bending technique [2], and makes use of multi-layer cantilever beam structures. Thin films for which elastic constants are to be determined (LPCVD SiN$_x$ in this paper) are deposited on well characterized films, which can sustain large elastic deformations. Thermal SiO$_2$ ideal for this application [3]. The elastic parameters of the top layer can be evaluated without the yielding of the bottom reference film . Furthermore, the bottom layers can withstand the large stresses that develop in the beams due to significant growth or thermal mismatches between the top and bottom films. This allows the residual stresses to be calculated with a simple elastic model.

After the fabrication of double-layer cantilever beams, the elastic constants of the nitride films are calculated from the force-deflection slope and known elastic constant of the thermal SiO$_2$. The curvature of this double-layer beam provides information on the residual stresses in the films. This test method is useful for measuring films that are either too thin (less than 0.5 µm thick) to test as single layer beams or for films that are easily attacked by the silicon etchant and therefore cannot be fabricated into single-layer beams.

SAMPLE PREPARATION PROCEDURES

A layer of wet-thermal SiO$_2$ was grown on a lightly doped P-type (100)Si wafer at 1100°C for about 5 hours. The thickness of the SiO$_2$ layer was measured at 1,465 µm using an ellipsometer. Two types of LPCVD nitrides, with different ratios of reactant gas flow rates, were deposited on the SiO$_2$ layer by the chemical reaction of SiCl$_2$H$_2$ and NH$_3$ at 785°C. The chamber pressure was 360 mtorr. There were no anneals after the depositions. In this paper, the nitride deposited with a gas flow ratio between SiCl$_2$H$_2$ and NH$_3$ of 1:3, will be refered to as a 'standard nitride'; whereas the nitride deposited with the ratio of 6:1 will be refered to as a 'low stress nitride.' The thicknesses of the standard and low stress nitride were 2140 ± 15 Å and 2870 ± 15 Å, respectively. The thickness measurements were made with both a surface profiler, which directly measures the step height of the nitrides on the oxides, and an ellipsometer, which measures the reference samples of nitrides on bare silicon wafers.

Figure 1: LPCVD SiN$_x$ (2100 Å)/SiO$_2$ (1.465 μm) beams.

Figure 2: Cross-sectional view of LPCVD SiN$_x$ (2100 Å)/SiO$_2$ (1.465 μm) beams.

The SiN$_x$ layer was patterned by standard photolithography processes and a dry etching process in a gas mixture of SF$_6$ and CF$_3$Br. The underlying SiO$_2$ layer was wet etched in a buffered hydrofluoric acid (BHF) solution. Finally, the exposed silicon was etched in an EDP solution (66 w/o ethylenediamine, 23 w/o DI water, 11 w/o pyrocatechol) at 115°C for 60 minutes, and then rinsed and dried. In this solution, the etch rates of thermal SiO$_2$, standard SiN$_x$, and low stress SiN$_x$ are extremely low (~ 00 Å/hr, ~35 Å/hr and ~45 Å/hr, respectively) compared to the etch rate of the Si ~60 μm/hr.

For the determination of average residual stress in the nitride, various thicknesses of the nitride films were produced using the dry etching process before the etching processes described above were initiated. The curvatures of the double-layer beams with different nitride thickness were measured in SEM photographs.

Figure 1 and Figure 2 are SEM photographs of multi-layer (standard SiN$_x$/SiO$_2$) cantilever beams which were fabricated using these procedures. All of these beams curve upwards due to initial compression in SiO$_2$ layer and tension in SiN$_x$ layer. As seen in these figures, the curvatures are smooth and circular in shape.

DETERMINATION OF ELASTIC CONSTANTS FOR VERY THIN FILMS

1. Mechanics for Double-layer Beam Bending

From the assumption that bending strain due to a curvature is continuous throughout the beam thickness, force and moment can be written in a general expressions as follows:

$$F = \int \sigma\, dZ = \varepsilon_0 \int E\, dZ + K \int EZ\, dZ \qquad (1)$$

$$M = \int \sigma Z\, dZ = \varepsilon_0 \int EZ\, dZ + K \int EZ^2\, dZ \qquad (2)$$

where σ is the normal stress, ε_0 is the axial strain, Z is the distance in a vertical direction from the neutral axis, and K is the curvature generated when a beam is deflected. The position of the neutral axis from the middle of double layer beam, z_0 (see Figure 3) can be determined by setting the coupling terms ($\int EZ\, dZ$) to zero. Thus
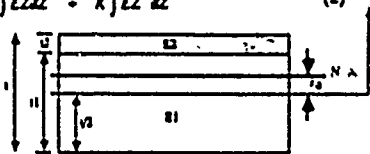

Figure 3: Schematic diagram of double-layer beam cross-section

$$r_0 = \frac{\int_{-t/2}^{t/2} E t \, dt}{\int_{-t/2}^{t/2} E \, dt} \tag{3}$$

Once the position of the neutral axis is known, the effective stiffness of a double-layer beam per unit width, $(EI)_{effective}$, can be calculated from eq. (2) as follows:

$$(EI)_{effective} = \int_{-t/2}^{t/2} E \{ t - t_0 \}^2 \, dt \tag{4}$$

$$= \int_{-t/2}^{(t_1 - t)/2} E_1 \{ t - t_0 \}^2 \, dt + \int_{(t_1 - t)/2}^{t/2} E_2 \{ t - t_0 \}^2 \, dt \tag{5}$$

Equations (3) and (5) suggest that t_0 and $(EI)_{effective}$ are functions of the elastic constants and the thicknesses of a multi-layer beam, e.g. E_1, E_2, t_1, t_2.

2. Test Procedures and Data

After the samples were prepared in double-layer cantilever structures, bending tests were performed as described elsewhere[2]. Figure 4 shows data for the bending test of a double-layer cantilever beam which consists of standard SiN_x (2100 Å) and thermal SiO_2 (1.465 μm) layers. The width and effective length of the beam were 22.5 μm and 33.1 μm, respectively. The Young's modulus of thermal SiO_2, which is the bottom layer in this sample, was measured earlier to be ~65 GPa [2]. The slope of the initial straight line in Figure 4 represents the spring constant of the Nanoindenter. When the tip of the indenter hits the beam and starts to deflect it the slope changes abruptly. The slope in the second linear section represents the sum of the spring constants of the machine and the beam. From this slope, it is possible to extract the actual effective stiffness of the beam. This measured value of $(EI)_{effective}$ will be used to determine the elastic constant of the second layer material, which is standard SiN_x in this example.

Since we know all of the physical dimensions of the double-layer beam and the Young's modulus of the bottom layer, the position of the neutral axis can be determined using eq. (3), if any arbitrary number is given for the Young's modulus of the second layer. The calculated position of the neutral axis in turn, determines the effective stiffness of the beam. Therefore, a set of values for $(EI)_{effective}$ for the beam can be calculated by varying the ratio of the elastic moduli of the first and second layers, E_2/E_1. Finally, the Young's modulus of the second layer is determined by comparing the measured stiffness with these calculated values: the correct ratio of E_2/E_1 makes the difference between the two stiffnesses vanish. As an example in Figure 5, the difference between the two stiffnesses for the same beam was plotted with the ratio, E_2/E_1, varying from 1 to 6. The difference goes to zero when the ratio is ~4.2. Thus, the Young's modulus of the standard
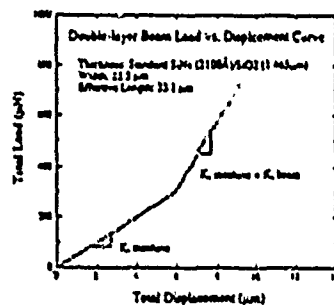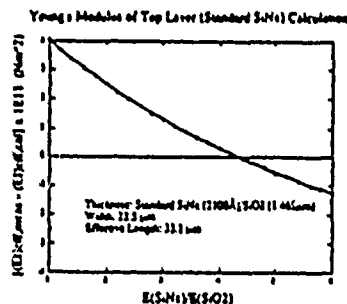

Figure 4: Deflection of a LPCVD SiN_x/SiO_2 beam.

Figure 5: Plot of difference in measured and calculated stiffnesses with E_2/E_1.

nitride is ~ 275 GPa, which is ~4.2 times greater than that of thermal SiO_2. Following the same test procedures and analysis, the Young's modulus of the low stress LPCVD nitrides was determined to be ~ 235 GPa.

Large deflections of the SiN_x/SiO_2 beams, to about 1/4 of their lengths, were also made. The normal stress at the surface section due to this large deflection is approximately 3 GPa (~ 1 % elastic strain). The total normal stress, including the residual stress to be discussed in next section, could be as high as ~ 4 GPa for the standard LPCVD nitride. However, we did not observe any fracture of the nitrides. This suggests that the films, up to the thicknesses we studied, have a very small average flaw size.

DETERMINATION OF RESIDUAL STRESSES FOR VERY THIN FILMS

1. Mechanics for Residual Stress Calculation from A Curvature

In order to calculate the average residual stress in thin films from the curvature of a multi-layer beam, we have adopted the elastic model proposed by P. H. Townsend *et al.* [4]. In their model, the curvature, K, and the normal stresses in each film, σ_i, are equated in terms of physical dimension, t_i, material constants, E_i, and hypothetical relaxed linear dimensions, d_i. If we modify this model for 2-dimensional cases with the coordinates used in Figure 3, the curvature and the normal stress can be rewritten as follows:

$$K = \frac{\sum_{i=1}^{\bullet}\frac{E_i}{(1-v_i)}\gamma_i\frac{t_i}{2}\left[-\ln d_i + \frac{\sum_{j=1}^{\bullet}\{E_i/(1-v_i)\}t_i\ln d_i}{\sum_{j=1}^{\bullet}\{E_i/(1-v_i)\}t_i}\right]}{\sum_{i=1}^{\bullet}\frac{E_i t_i}{(1-v_i)}\left[\left\{\frac{1}{2}(\frac{t_i}{2}+t_0)-\frac{t^2}{3}\right\}+(\frac{t_i}{2}-t_0)\frac{\gamma_i}{2}-\frac{1}{12}(3\gamma_i^2+t_i^2+t^2)\right]} \quad (6)$$

and

$$\sigma_i = \frac{E_i}{(1-v_i)}\left[-\ln d_i + \frac{\sum_{j=1}^{\bullet}\frac{E_i}{(1-v_i)}t_i\ln d_j}{\sum_{j=1}^{\bullet}\frac{E_i}{(1-v_i)}t_i}-(t_0-t)K\right] \quad (7)$$

where t is total thickness, and the coefficients γ_i and β_{ij} are defined as:

$$\gamma_i = \sum_{j=1}^{\bullet}\beta_{ij}t_j \quad (8)$$

and

$$\beta_{ij} = -1 \text{ (if } i > j), \ 0 \text{ (if } i = j), \ +1 \text{ (if } i < j). \quad (9)$$

Since we know all of the parameters (E_i, t_i), it is possible to calculate stresses in each film by measuring the ratio of the linear dimensions from curvatures. For the 3-layer case, as in the $SiN_x/SiO_2/Si$ system discussed in next section, it is sufficient to measure two ratios, (d_2/d_1) and (d_3/d_2), for the stress calculation.

2. Calculation Results and Discussion

Using the elastic model discussed above, the residual stresses in LPCVD nitrides, deposited on thermally oxidized silicon substrate, were calculated for various nitride film thicknesses.
In order to determine (d_{SiO_2}/d_{Si}) and the average stress in the SiO_2, a laser scanning technique was used. The stress was determined to be ~ -280 MPa. We know that the stress is uniformly distributed throughout the film thickness since single-layer SiO_2 beams with various thicknesses show no curvature. For the determination of (d_{SiN_x}/d_{SiO_2}), the curvatures of SiN_x/SiO_2 beams were measured from SEM photos. We have used 180.5 GPa for $(E/(1-v))_{Si}$ [5], 0.16 [6] and 0.2

for the Poisson's ratios of SiO_2 and SiN_x, throughout the calculations.

The radii of SiN_x/SiO_2 beams are plotted in Figure 6 for various film thicknesses. This plot shows that the radius of the standard SiN_x/SiO_2 beam is smaller than the one of low stress SiN_x/SiO_2 beam at the same thickness. Therefore, the residual stress in the standard nitride is higher, see Figures 7 and 8. For relatively thick films (>1200 Å), the average stresses in the standard and low stress nitrides are ~1.15 GPa and ~440 MPa, respectively. The difference in residual stress can be explained at least partially by the difference in silicon contents in nitrides. The refractive index, N_f, of the standard nitride (2140 Å) was measured at 2.00, which is an accepted value for stoichometric Si_3N_4 [7]. On the other hand, N_f of the low stress nitride (2870 Å) was measured to be ~2.25, for various SiN_x thicknesses.

Figure 6: Plot of radius of SiN_x/SiO_2 beams for various SiN_x thicknesses.

Even though this technique does not directly quantify the composition, it indicates that the silicon content in low stress nitride is higher than in stoichometric films. Many reports in the literature have shown that the excess silicon relieves tension in nitride films [for example, see 8]. We have also found that the refractive index was not a constant with the film thickness in both nitrides. This suggests that these films are not completely homogenious.

Figures 7 and 8 show the average stresses, σ_{avg}, and the differential stresses, σ_d, for both nitrides at room temperature. The differential stress between two film thicknesses, T_{n-1} and T_n, was defined as follow:

$$\sigma_d(T_{n-1}, T_n) = \frac{\sigma_{avg, T_n} T_n - \sigma_{avg, T_{n-1}} T_{n-1}}{T_n - T_{n-1}} \tag{10}$$

As seen in these figures, the average stresses in both nitrides increase with thickness when the films are very thin, and that they approach constant values (~1.15 GPa and ~440 MPa) after certain thicknesses. The main source of error source in these calculations is the uncertainty in film thickness. We assumed that the nonuniformity in film thickness, after all etching processes, is ≈ 50Å. This was accounted for by the error bars at each data point. Since the portion of the error range with respect to the total thickness increases as the film gets thinner, the magnitude of the error bar also increases accordingly. However, the general trend can still be clearly seen. The accuracy of this technique largely depends on two main factors: the thickness measurement of top film, and the initial thickness of the bottom layer. If we use a thin bottom layer and accurately measure the top layer thickness, it is possible to detect very small residual stresses.
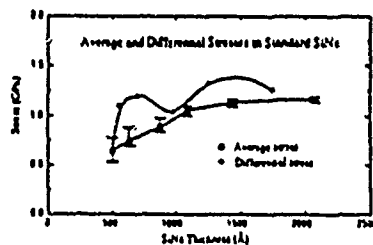

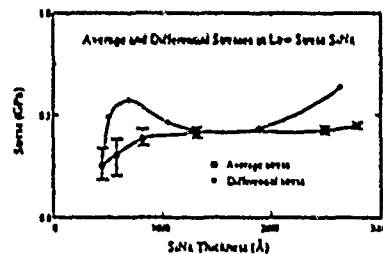Figure 7: Average and differential stresses in standard nitride film.

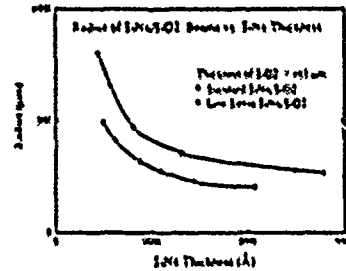Figure 8: Average and differential stresses in low stress nitride film.

For very thin films, the relatively large change of the average stress indicates the variation of the differential stress. The differential stress increases gradually with a small slope for the regions where the average stress starts to approach a constant value. The trend of the actual stress states was verified by removing the bottom oxide layer with BHF solution, thus making a single-layer nitride beam. Figure 9 shows the cross sectional view of the nitride beams. The curvature of the nitride beams is caused by the non-uniform distribution of the actual stress in the film, and also indicates clearly that, in an average sense, more tension develops in the top portion of the film. However, it is not understood well, at this point, why the variations of the normal stresses are present in these films. One possible explanation is that silicon rich SiN_x grows near the interface, and then the composition approaches the proper stoichiometry (a steady value in low stress nitride case) as the film grows thicker.
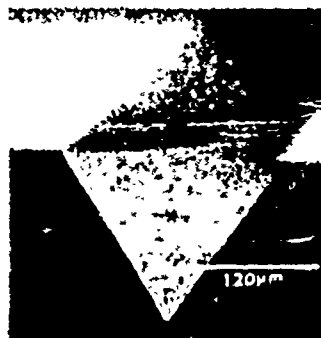
Figure 9: Cross-sectional view of standard SiN_x (2100 Å) beams.

SUMMARY

A unique technique which can measure both the elastic parameters and residual stresses of very thin films has been developed using multi-layer micro-cantilever beam structures. A film system of LPCVD $SiN_x/SiO_2/Si$ was tested using the technique and showed that:

(1) The Young's moduli of LPCVD standard and low stress SiN_x, and PECVD SiN_x, were measured at 275 and 235 GPa, respectively.

(2) The average stress at room temperature in both LPCVD standard and low stress SiN_x increases near the interface between SiN_x/SiO_2 and approaches constant values, 1.15 GPa and 440 MPa, at larger thicknesses.

ACKNOWLEDGEMENTS

REFERENCES

[1] A. Bohg and A. K. Gained, Appl. Phys. Lett. 33, pp 895-897, Nov., 1978.
[2] T. P. Weihs, S. Hong, J. C. Bravman and W. D. Nix, J. Mater. Res. 3, pp 931-942, Sep/Oct., 1988.
[3] T. P. Weihs, S. Hong, J. C. Bravman and W. D. Nix, These Proceedings.
[4] P. H. Townsend, D. M. Barnett and T. A. Brunner, J. Appl. Phys. 62, pp. 4438-4444, Nov., 1987.
[5] W. A. Brantley, J. Appl. Phys. 44, pp 534-535, Jan., 1973.
[6] CRC Handbook of Chemistry and Physics, edited by R. C. Weast (CRC, Boca Raton, FL, 1985), p.56
[7] K. E. Bean, P. S. Gleim, R. L. Yeakley and W. R. Runyan, J. Electrochem. Soc., 114, pp. 733-737, July, 1967.
[8] P. Pan and W. Berry, J. Electrochem. Soc., 132, pp. 3001-3005, Dec., 1985.

ANALYSIS OF ELASTIC AND PLASTIC DEFORMATION ASSOCIATED WITH INDENTATION TESTING OF THIN FILMS ON SUBSTRATES

A. K. BHATTACHARYA

Department of Mechanical Engineering, Stanford University, Stanford, CA 94305, U.S.A.

and

W. D. NIX

Department of Materials Science and Engineering, Stanford University, Stanford,
CA 94305, U.S.A.

Abstract—A study has been made of the elastic and plastic deformation associated with sub-micrometer indentation of thin films on substrates using the finite element method. The effects of the elastic and plastic properties of both the film and substrate on the hardness of the film substrate composite are studied by determining the average pressure under the indenter as a function of the indentation depth. Calculations have been made for film substrate combinations for which the substrate is either harder or softer than the film and for combinations for which the substrate is either stiffer or more compliant than the film. It is found, as expected, that the hardness increases with indentation depth when either the yield strength or the elastic modulus of the substrate is higher than that of the film. Correspondingly, the hardness decreases with indentation depth when the yield strength or elastic modulus of the substrate is lower than that of the film. Functional equations have been developed to predict the hardness variation with depth under these different conditions. Finite element simulation of the unloading portion of the load-displacement curve permits a determination of the elastic compliance of the film substrate composite as a function of indentation depth. The elastic properties of the film can be separated from those of the substrate using this information. The results are in good agreement with King's analytical treatment of this problem.

INTRODUCTION

Knowledge of the strengths of metallic thin films bonded to substrates is becoming more and more important due to the increased use of various metallization techniques in integrated circuit devices. This is important because an understanding of the mechanics of deformation permits one to better predict the conditions leading to failure in these thin film structures. This knowledge also permits better design from a structural point of view. The sub-micrometer indentation technique is emerging as one of the popular ways to study the elastic and plastic deformation of thin films on substrates (Pethica *et al.*, 1983; Doerner and Nix, 1986; Doerner *et al.*, 1986). This technique is particularly attractive because it provides information about elastic and plastic deformation on a highly localized scale and because it is simple to use. In the present study we describe the elastic and plastic deformation associated with the indentation of thin films on substrates from a continuum point of view. We have used the finite element simulation technique to study the effects of the yield strength and elastic modulus of the film and substrate on the hardness and elastic compliance of the film/substrate composite.

Although indentation tests have been used for a long time, few theoretical treatments of indentation deformation are available. Such treatments are required to obtain fundamental information about the elastic and plastic properties of the film from an indentation experiment. For an introduction to some of the published literature on this subject, the interested reader may refer to a recent paper (Bhattacharya and Nix, 1988). In this work we concentrate on small indentations of the kind produced in typical sub-micrometer indentation tests used for thin films on substrates. The hardnesses of several thin film materials using sub-micrometer indentation tests have been measured (Doerner and Nix, 1986). The elastic properties of these films were also determined as a function of film thickness and an

empirical equation to describe the effect of the substrate on the compliance of the film-substrate composite was suggested. This problem was later examined analytically by King (1987), who developed a theoretical analysis for determining the compliance. While it is possible to study elastic indentation of thin films using analytical methods, elastic-plastic indentations are sufficiently complex that finite element techniques are required. A finite element study of indentation of bulk materials has been conducted recently by the present authors (Bhattacharya and Nix, 1988). We showed that the basic features of the indentation experiment can be described using the finite element technique in conjunction with relatively simple material behavior. In this paper we extend the analysis to the problem of thin films on substrates. Here we analyze the effect of the properties of both the film and the substrate on the hardness and the elastic compliance of the film/substrate composite.

THE FINITE ELEMENT MODEL

Sub-micrometer indentation testing permits the measurement of force-distance relations on a very small scale; a detailed description of this can be found elsewhere (Pethica et al., 1983; Doerner and Nix, 1986). Simulations of these force-distance relations for the indentation of thin films on substrates using a rigid indenter were performed using the large strain elasto-plastic feature of the ABAQUS (1985) finite element code, with tensile stress-strain data as input. Film and substrate materials with various yield strengths and Young's moduli were studied. The quasi-static nature of the process permits us to use the static analysis performed by the program. Underlying the approach in this code is the discretization of the continuum involved (the layer to be indented here); the indenter was considered to be perfectly rigid. Also, an important feature of this program involves the capability to model contact between the indenter and the sample as a sliding interface. The initial nodal gaps between the indenter and the surface of the specimen were prescribed; the program automatically keeps track of their change and indicates any gap closure or opening in a particular specified direction. These interface elements thus simulate contact between the indenter and the specimen surface. Whenever the closure distance between the indenter and the specimen becomes zero, contact is assumed and an external reaction force is exerted on that particular material point to keep it moving along with the indenter. Because the program calls for incremental loading and also makes use of interface elements, the expanding contact area associated with indentation occurs naturally whenever new interface elements come into contact.

In this analysis, the indenter and specimen are treated as bodies of revolution to avoid the inherent three-dimensional nature of the problem of indentation with a pyramid shaped indenter (Pethica et al., 1983; Doerner and Nix, 1986). This approximation is considered to be acceptable for the case of continuum plasticity; a three-dimensional analysis would be needed to treat crystal plasticity. In the present treatment the pyramid indenter was approximated by an axisymmetric cone (having a perfectly sharp tip) of equal volume for a given indenter depth. In actual test conditions, the indenter tip has a finite radius, thus giving rise to a somewhat different response. The specimen consists of a thin film (1 μm thick) on a semi-infinite substrate approximated as a plate two orders of magnitude thicker than the film. Perfect contact between the film and the substrate is assumed and both film and substrate were considered to be initially stress free. The indenter and specimen are shown schematically in Fig. 1, along with the appropriate boundary conditions for the problem. Symmetry properties have been used to simplify the boundary conditions. During preliminary simulations the boundary condition on the surface on the right-hand side of the specimen was changed from fixed radial displacements to traction free; this change had no effect on the indentation parameters, thus showing that this boundary was indeed remote. Because very small indentations were being simulated, the meshes near the indenter needed to be very fine to be able to describe the deformation and stress gradients associated with indentation with sufficient accuracy. Thus, extremely fine mesh sizes were used under the indenter; they became progressively coarser at distances further away from the indenter. Axisymmetric four noded elements were used for the continuum. In order to obtain an accurate estimate of the radius of the contact area, an extremely fine mesh thickness of the
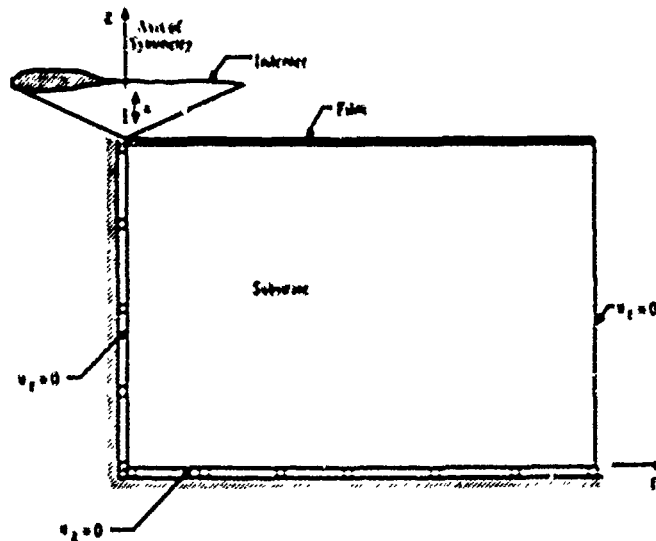
Fig. 1. Schematic diagram of the specimen under the indenter showing the boundary conditions of the indentation problem

order of 0.02 μm had to be used along most of the indenter contact surface. To keep the required computer time within limits, a total of 461 elements including the interface elements were used to represent the deformed material. Figure 2 shows a magnified view of the elements near the indenter and the staircase arrangement for the other elements at points further away from the indenter.

To simulate a typical indentation process, a downward displacement (negative z-direction in Fig. 1) was imposed on the indenter; this causes the indenter to push into the surface of the material. At the end of the indentation experiment the indenter was given an upward displacement until it was free of contact with the specimen. For a given indenter displacement, the corresponding load was determined by summing the reaction forces at the contact node points on the indenter. The interface between the specimen and the indenter was assumed to be frictionless since no noticeable change in the load-displacement response was observed by using a friction coefficient of 1. The mesh thickness of 0.02 μm along the indenter contact surface was determined to be acceptable by finding the mesh size below which no further significant changes in the indentation load-displacement response were observed.
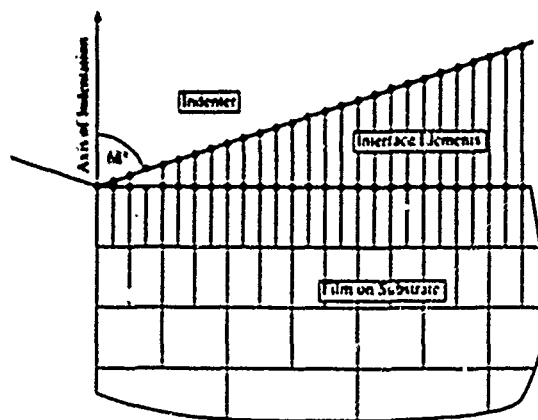

Fig. 2. Detailed pattern of the mesh distribution near the indenter showing the interface elements.

Table 1. Yield strengths and Young's moduli used for the finite element simulation of indentation of thin films on substrates ($\sigma_y/\sigma_s = 0.2$ s)

Yield strength (MPa)						
σ_f	485	485	2070	4410	2070	4410
σ_s	4410	2070	4410	485	485	2070
σ_f/σ_s	0.11	0.233	0.47	9.13	4.39	2.13
Young's modulus (GPa)						
E_f	46.5	76	127	207	345	
E_s	127	1237	127	122	127	
E_f/E_s	0.367	0.6	1.0	1.67	2.73	

Table 2. Elastic and plastic properties of aluminum and silicon used in the finite element analysis

Material	Young's modulus (GPa)	Poisson's ratio	Yield strength (MPa)
Aluminum	75.9	0.33	485
Silicon	127	0.278	4410

The constitutive model for the specimen material (both film and substrate) was that of an elastic-plastic von Mises material. Only the case of no strain hardening (i.e. both materials assumed to be elastic-fully plastic) was considered. The elastic and plastic properties of the films and substrates used in the various calculations are given in Tables 1 and 2. The finite element calculations were performed using an IBM 4341 mainframe computer with run times of 1-2 days and also using a Vax II Workstation with run times of 5-6 days for average indentation depths.

RESULTS AND DISCUSSION

In our previous work (Bhattacharya and Nix, 1988), we showed that the finite element analysis can adequately describe the load-displacement response observed in a typical submicrometer indentation test. We also showed how the hardness and Young's modulus can be evaluated from such a simulated result. Figure 3 shows a simulation result for indentation of silicon together with an experimental result reported by Pethica et al. (1983). There are some differences in the two results that can be attributed to differences in the actual and assumed material properties and to the discontinuous nature of the simulation procedure.

When an indentation test is performed on a single homogeneous material, the hardness remains essentially constant, irrespective of the depth of indentation. This result is expected
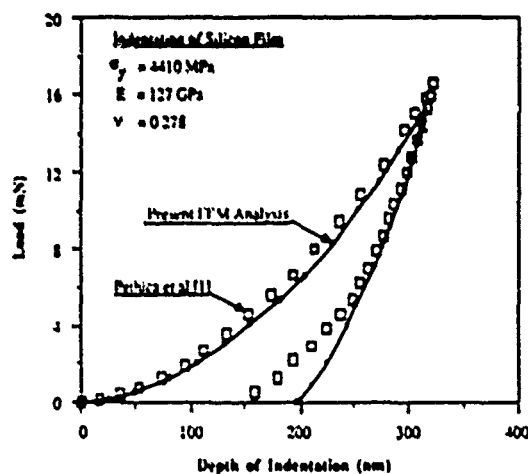

Fig. 3. Comparison of the FEM results of a previous paper (Bhattacharya and Nix, 1988) with the experimental results of Pethica et al. (1983) for indentation of silicon.
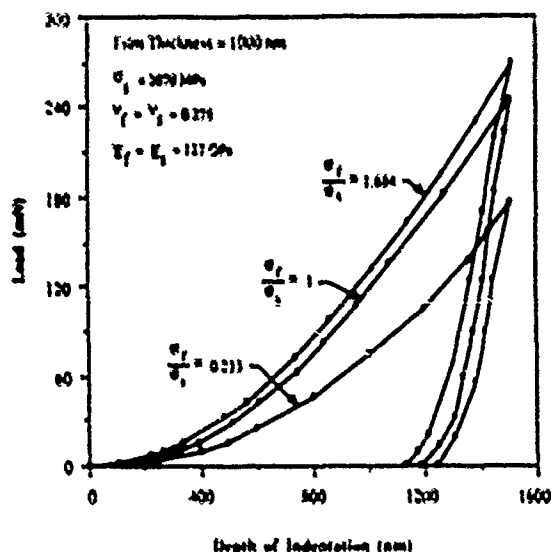
Fig. 4. Effect of relative yield strengths on the load-depth response for thin films on substrates.

in a continuum treatment. Johnson (1970) developed a theoretical expression for hardness using an expanding spherical cavity model of the indentation process. He found that the hardness depends only on the elastic and plastic properties of the material in question and not on the depth of indentation. This result was also found in our finite element study of indentation. In contrast to this behavior, if an indentation is made on a thin film attached to a substrate having different elastic and plastic properties, the hardness associated with the indentation process is expected to continually change due to the gradually increasing influence of the substrate. We next present various results for the indentation of thin films on substrates.

Hardness

One of the goals of this work has been to understand how the hardness of a thin film changes with depth of penetration of the indenter. The most important quantities to be considered in this analysis are the thickness of the film and the yield strength, Young's modulus and Poisson's ratio of both the film and the substrate. Strain hardening in the film and substrate could also be considered but it has not been included in the present analysis. In our previous work on the indentation of semi-infinite solids we found that including strain hardening in the analysis does not produce qualitatively significant effects. The response of a material with a high rate of strain hardening is essentially the same as the response of a material with a higher yield strength. In a typical indentation experiment, the elastic-plastic response is characterized by a growing volume of material subjected to a fixed strain rather than by a fixed volume of material subjected to an increasing strain. For this reason it is not especially important to include strain hardening in the analysis.

We consider first the hypothetical cases in which the film and substrate have either different elastic properties (Young's moduli) or different yield strengths. As mentioned above, all simulations have been performed on 1 μm thick films that are perfectly bonded to a comparatively thick substrate. It is shown below that even though we have used only one film thickness, the formulation developed from these results will also predict the general trends observed for films with different thicknesses. Figure 4 shows the calculated load-displacement response in cases for which the yield strengths of the film and substrate are different but for which both Young's moduli and Poisson's ratio for the film and the substrate are the same. Similarly, Fig. 5 shows the response in cases for which the Young's moduli of the film and substrate are different but for which the yield strengths and Poisson's ratios are the same. From these results, the hardness is calculated as the load divided by
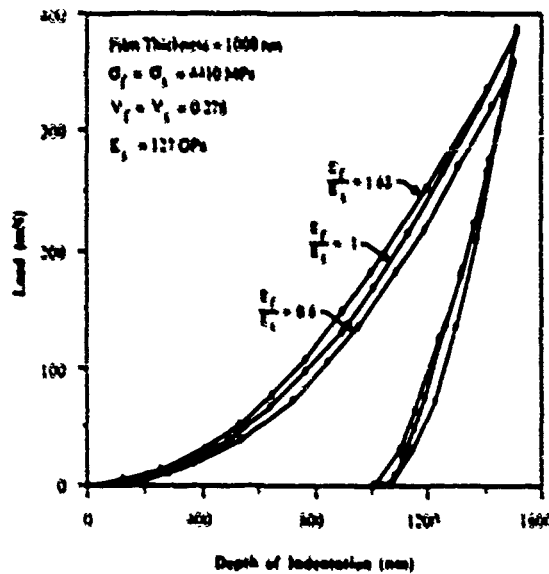
Fig. 5. Effect of relative Young's moduli on the load-depth response for thin films on substrates.

the projected area of contact. These hardness results are shown in Figs 6 and 7. The calculated hardness and the depth of indentation have been non-dimensionalized with the hardness of the substrate and the film thickness, respectively. As intuitively expected, for the case of a softer film on a harder substrate (Fig. 6), the hardness increases with the indentation depth. As seen for this case, the hardness is independent of the substrate for indentation depths less than about 0.3 of the film thickness, after which the hardness slowly increases because of the presence of the substrate. For the case of a harder film on a softer substrate (Fig. 7), the hardness decreases as the depth of indentation increases. In this case, the hardness is constant for indentation depths less than about 0.2 of the film thickness. The regime of constant hardness appears to get smaller as the yield strength of the film increases relative to that for the substrate.

In Figs 8 and 9, we show the hardness results for cases in which the film and substrate have different Young's moduli. It is observed that the variation of hardness with depth of
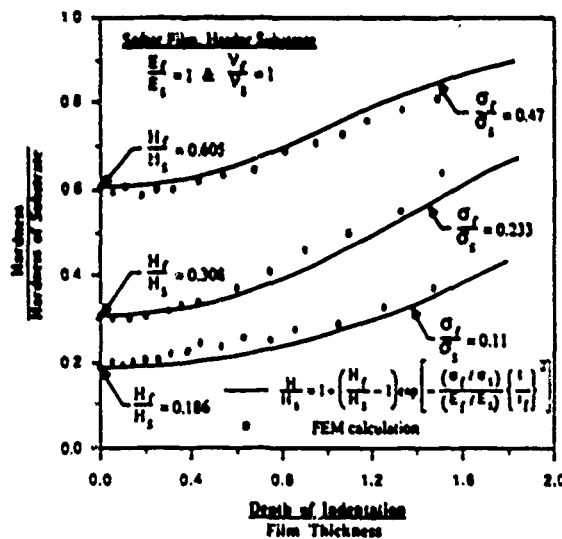

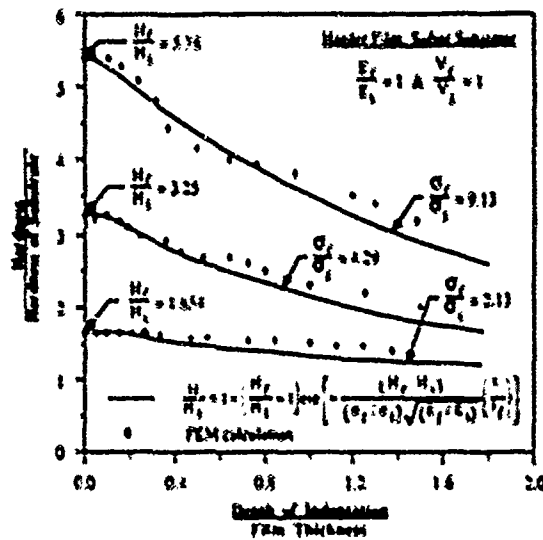Fig. 6. Effect of relative yield strengths on the hardness of soft films on hard substrates.

Fig. 7 Effect of relative yield strengths on the hardness of hard films on soft substrate.

indentation in these cases is qualitatively similar to cases in which the film and substrate have different yield strengths, although the hardness changes more gradually than in the previous cases.

The results presented in Figs 6–9 can be used to develop general relations for the hardness variation with indentation depth. It is believed that the hardness results will depend only weakly on Poisson's ratio. For this reason, the effect of Poisson's ratio has not been considered in our analysis. After analyzing these results and fitting with various forms of equations, we have developed two empirical equations which satisfactorily describe the variation of hardness with depth. For the case of a soft film on a harder substrate, the effect of the substrate on film hardness can be described as

$$\frac{H}{H_s} = 1 + \left(\frac{H_f}{H_s} - 1\right)\exp\left[-\frac{(\sigma_f/\sigma_s)}{(E_f/E_s)}\left\{\frac{t}{t_f}\right\}^2\right]$$

(1)

where, E_f, E_s are the Young's moduli, σ_f, σ_s the yield strengths and H_f, H_s the hardnesses

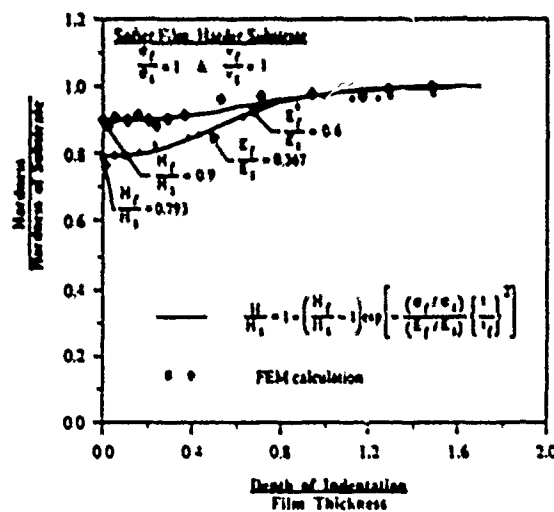
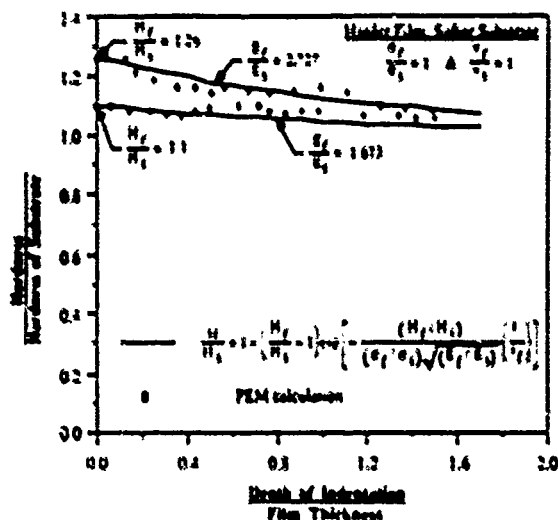Fig. 8. Effect of relative Young's moduli on the hardness of soft films on hard substrates.

Fig. 9. Effect of relative Young's moduli on the hardness of hard films on soft substrates.

of the film and substrate, respectively. Similarly for the case of a hard film on a softer substrate, the hardness can be expressed as

$$\frac{H}{H_s} = 1 + \left(\frac{H_f}{H_s} - 1\right)\exp\left[-\frac{(H_f/H_s)}{(\sigma_f/\sigma_s)\sqrt{(E_f/E_s)}}\left\{\frac{t}{t_f}\right\}\right].$$ (2)

These two equations are fully nondimensionalized and thus can be used for any film thickness involving any material combination of the film and substrate.

We next discuss the predictive capabilities of these functional equations to describe the hardness of an aluminum film on a silicon substrate and a silicon film on an aluminum substrate. Figure 10 compares the load-displacement response for the indentation of an aluminum film on a silicon substrate with that for a silicon film on an aluminum substrate. The elastic and plastic properties for these materials used in the FEM calculation are given in Table 2. In Fig. 10, we also show the indentation curves for bulk aluminum and silicon. At very small depths of indentation, the load-depth response for the film/substrate
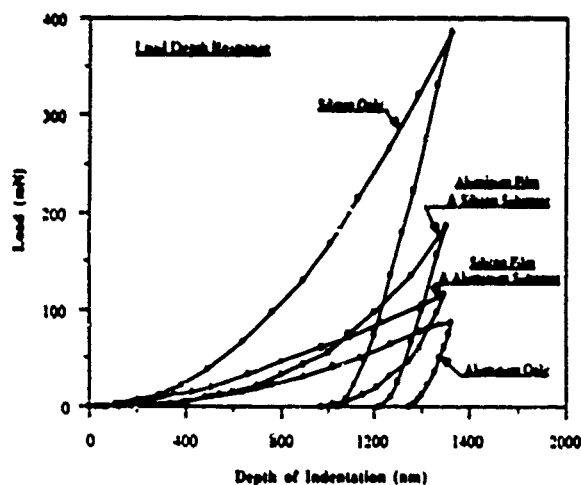

Fig. 10. Load-depth responses for the indentation of an aluminum film on a silicon substrate and a silicon film on an aluminum substrate.
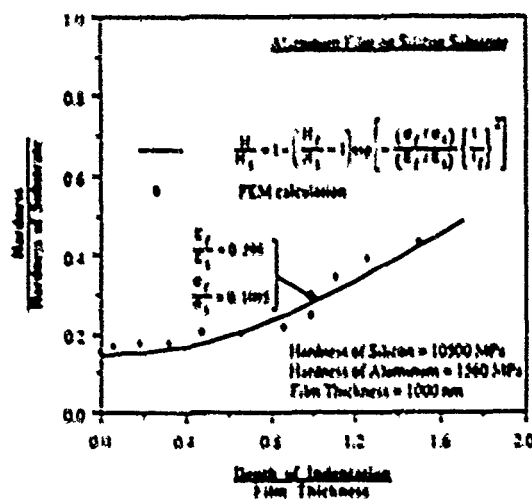
$$\frac{H}{H_s} = 1 + \left(\frac{H_f}{H_s} - 1\right) \exp\left[-\frac{(\sigma_f/\sigma_s)}{(E_f/E_s)}\left(\frac{t}{T_f}\right)\right]$$

FEM calculation

Hardness of Silicon = 10900 MPa
Hardness of Aluminum = 1560 MPa
Film Thickness = 1000 nm

Fig. 11 Comparison of hardness values predicted by eqn (1) with FEM calculations for the case of an aluminum film on a silicon substrate

composite is the same as the response for the thin film material in bulk form. This result is as expected since for very small indentations only the properties of the thin film are involved. Figure 11 shows the hardness values obtained from Fig. 10 plotted as a function of indentation depth for the case of an aluminum film on a silicon substrate. Similarly, hardness values for a silicon film on an aluminum substrate are plotted in Fig. 12. We also show predictions of eqns (1) and (2), respectively, in these figures. We find that the empirical relations developed for hypothetical films and substrates give a good description for the specific cases of films and substrates composed of aluminum and silicon.

For the case of a soft film on a harder substrate, eqn (1) indicates that as the film is made thinner, it becomes progressively harder. On the other hand, in the case of a harder film on a softer substrate, eqn (2) predicts that as the film thickness is reduced, the hardness of the film decreases. These results suggest that the substrate can play a dominating role in increasing or decreasing the hardness of a film.
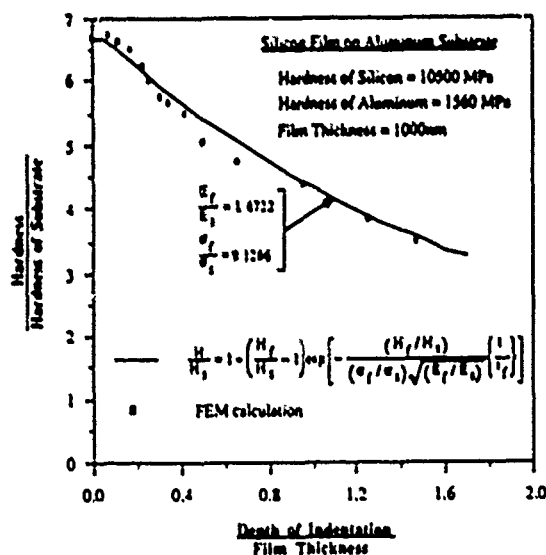

$$\frac{H}{H_s} = 1 + \left(\frac{H_f}{H_s} - 1\right) \exp\left[-\frac{(H_f/H_s)}{(\sigma_f/\sigma_s)\sqrt{(E_f/E_s)}}\left(\frac{t}{T_f}\right)\right]$$

FEM calculation

Silicon Film on Aluminum Substrate

Hardness of Silicon = 10900 MPa
Hardness of Aluminum = 1560 MPa
Film Thickness = 1000nm

Fig. 12. Comparison of hardness values predicted by eqn (2) with FEM calculations for the case of a silicon film on an aluminum substrate.

Elastic response

Recent experimental studies (Pethica *et al.*, 1983, Doerner and Nix, 1986), have shown that in an indentation experiment, the process of unloading is elastic. This is indicated by the linearity of the initial portion of the unloading curve. This has been confirmed by the present authors (Bhattacharya and Nix, 1988), through an FEM analysis of the indentation of silicon and aluminum. We also observed that during the earliest stage of unloading, the contact area under the indenter remains constant. When these conditions of unloading exist, Young's modulus of the material being indented can be determined from the compliance of the film and the plastic depth (Pethica *et al.*, 1983; Doerner and Nix, 1986). As demonstrated previously (Doerner and Nix, 1986), the plot of compliance vs reciprocal plastic depth is linear for a bulk material. However, in the case of thin films on substrates, this linear relationship no longer holds due to the varying contribution of the film and substrate to the measured compliance. There are two limiting cases. For indentation depths that are very small compared to the film thickness, the compliance approaches the value expected for the thin film material whereas for large depths of indentation, it approaches the value expected for the substrate material. The variation of overall compliance of a thin film/substrate composite was studied previously (Doerner and Nix, 1986). Later King (1987) used basis functions and an integral equation technique to perform elastic analyses for punches of different shapes and derived more accurate equations for the compliance in terms of the projected area of contact under the indenter. King's equation, when simplified for an infinitely rigid indenter, becomes

$$\frac{dh}{dP} = \frac{1}{\beta\sqrt{A}}\left[\frac{(1-v_f^2)}{E_f}\left\{1-\exp\left(-\frac{\alpha' t_f}{\sqrt{A}}\right)\right\} + \frac{(1-v_s^2)}{E_s}\left\{\exp\left(-\frac{\alpha' t_f}{\sqrt{A}}\right)\right\}\right] \qquad (3)$$

where A is the projected area of contact under the indenter, β a numerical factor related to the shape of the indenter and α' a parameter dependent on the depth and is obtainable from a set of curves given by King (1987).

In eqn (3), the value of A is related to the plastic depth through the indenter geometry and hence, $A = k^2 h_p^2$, where $k^2 = 24.5$ is a constant related to the geometry of the indenter. With this transformation and normalizing in terms of t_f, h_p, we can rewrite eqn (3) as

$$t_f\frac{dh}{dP} = \frac{1}{\beta k}\left(\frac{t_f}{h_p}\right)\left[\frac{(1-v_f^2)}{E_f}\left\{1-\exp\left(-\alpha''\frac{t_f}{h_p}\right)\right\} + \frac{(1-v_s^2)}{E_s}\left\{\exp\left(-\alpha''\frac{t_f}{h_p}\right)\right\}\right] \qquad (4)$$

where $\alpha'' = \alpha'/k$.

This equation is expected to be valid for any film thickness t_f. Compliances dh/dP were directly obtained from the unloading part of the FEM calculations based on a curve fitting procedure described in our earlier analysis (Bhattacharya and Nix, 1988). These compliances were determined for different combinations of indentation depths and film thickness. In Fig. 13, we have plotted $t_f dh/dP$ values for film/substrate composites having various Young's moduli ratios and compared them with the predictions of eqn (4). In this treatment the film and the substrate are considered to have the same Poisson's ratio. Calculated values from eqn (4) are seen to be in reasonable agreement with the FEM results for large depths of indentation. But, for smaller depths of indentation, it appears that the values predicted by this equation asymptotically approach the compliance of the film too slowly.

We next consider the elastic response of two specific film/substrate composites, e.g. an aluminum film on a silicon substrate and a silicon film on an aluminum substrate. These results, based on the material properties given in Table 2, are presented in Figs 14 and 15. In these plots, comparison has been made between the FEM calculations and the predictions of eqn (4). The two results are in reasonably good agreement.
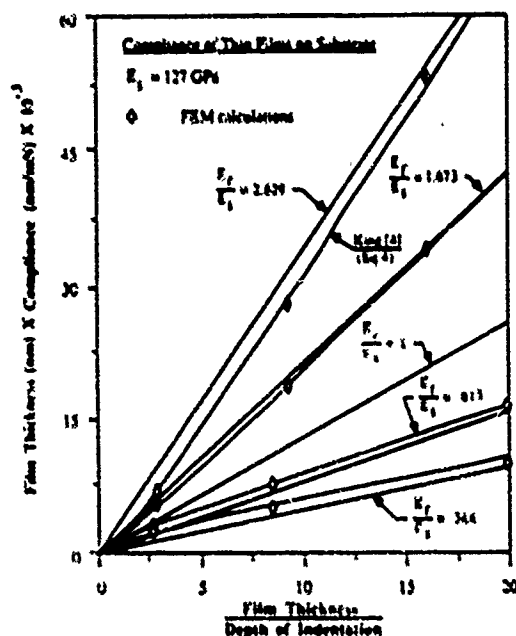
Fig. 13. Comparison of the compliances predicted by eqns (3) and (4) with the FEM calculations for various Young's moduli.

CONCLUSIONS

The results obtained in this study indicate that the finite element technique can be used to simulate the elastic and plastic response of a sub-micrometer indentation test for thin films on substrates. Based on calculations for two different classes of film/substrate composites, namely soft films on harder substrates and hard films on softer substrates, functional equations have been developed to describe the variation of hardness with depth of indentation. These equations correctly indicate that the film strength increases or decreases as the film thickness is reduced depending on whether the film is softer or harder, respectively,
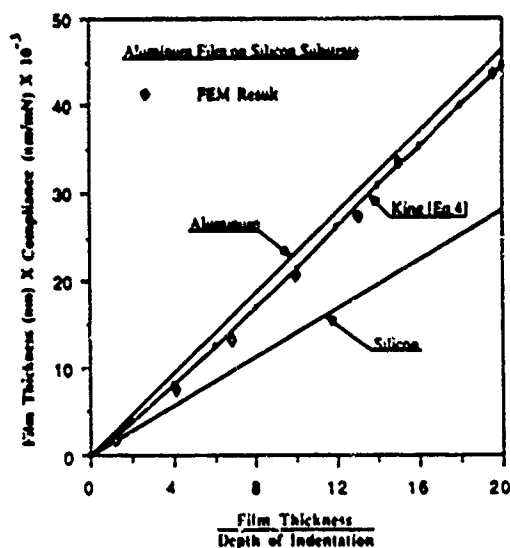

Fig. 14. Comparison of the FEM calculations with the predictions of eqn (4) for an aluminum film on a silicon substrate.
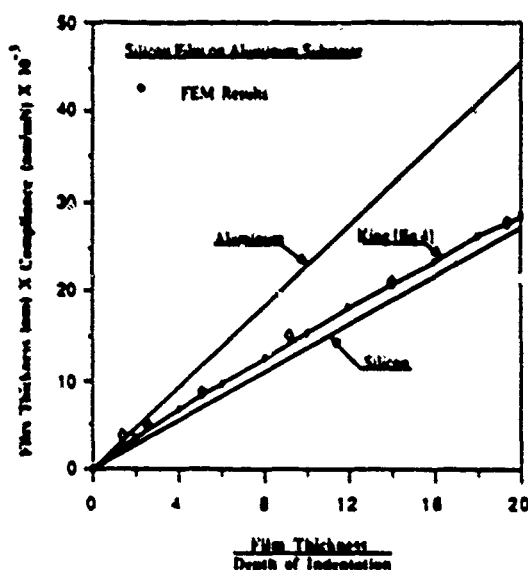
Fig. 15. Comparison of the FEM calculations with the predictions of eqn (4) for a silicon film on an aluminum substrate.

than the substrate. The effect of changing film thickness is found to be larger for the case of a soft film on a harder substrate. It is also shown that the elastic compliance of the film/substrate composite can be determined as a function of indentation depth from the unloading part of the load-displacement curves obtained from the FEM analyses. The results are in close agreement with King's analytical treatment of this problem.

Taken altogether, the results of this work provide a scheme by which the hardness of a very thin film can be determined from the measured hardness of a thin film/substrate composite. First the hardness and elastic modulus of the bare substrate would be measured by indentation methods. Then, the hardness and elastic compliance of the thin film/substrate composite would be measured as a function of indentation depth. The variation of hardness with depth would indicate whether eqn (1) or eqn (2) should be used to describe the properties of the composite. The compliance data would be used in conjunction with eqn (4) to determine the elastic modulus of the film or, equivalently, E_f/E_s. Finally, eqn (1) or eqn (2), as appropriate, would be inverted to obtain the hardness of the film from the hardness of the thin film/substrate composite.

REFERENCES

ABAQUS finite element program (1985). HKS Inc., Providence, Rhode Island.
Bhattacharya, A. K. and Nix, W. D. (1988). Finite element simulation of indentation experiments. *Int. J. Solids Structures* 24, 881-891 (1988).
Doerner, M. F., Gardner, D. S. and Nix, W D. (1986). Plastic properties of thin films on substrates as measured by sub-micron indentation hardness and substrate curvature techniques. *J. Mater Res.* 6, 845-851.
Doerner, M. F. and Nix, W. D. (1986). A method for interpreting the data from depth-sensing indentation instruments. *J. Mater. Res.* 4, 601-609.
Johnson, K. L. (1970). The correlation of indentation experiments. *J. Mech. Phys. Solids* 18, 115-126.
King, R. B. (1987). Elastic analysis of some punch problems for a layered medium. *Int. J. Solids Structures* 23, 1657-1664.
Pethica, J. B., Hutchings, R. and Oliver, W. C. (1983). Hardness measurements at penetration depths as small as 20 nm. *Phil. Mag. (A)* 48, 593-606.

FINITE ELEMENT SIMULATION OF INDENTATION
EXPERIMENTS

A. K. Bhattacharya
Department of Mechanical Engineering, Stanford University, Stanford,
CA 94305, U.S.A.

and

W. D. Nix
Department of Materials Science and Engineering, Stanford University, Stanford,
CA 94305, U.S.A.

Abstract—Indentation experiments are now being used to study the elastic and plastic properties
of materials on a sub-micrometer scale. Simulations of these experiments have been performed using
the finite element method under the conditions of frictionless and completely adhesive contact and
within the context of incremental elasto-plasticity. Comparison of these simulated results with
experimental results demonstrates that the continuum based finite element approach has the capa-
bility to determine the load-depth response of a sub-micrometer indentation test. It is also shown
that the hardness and elastic modulus of the material can be obtained from the loading and
unloading portions of these simulated curves.

INTRODUCTION

In recent years various sophisticated thin film deposition techniques have been developed
in response to demands for better materials for high technological applications, especially
in miniaturized electronic components. These techniques, such as: sputtering, vapor depo-
sition, ion implantation, laser glazing and other modern surface modification techniques,
involve tailored control of the mechanical and structural properties of the material involved.
With respect to mechanical properties, these kinds of control depend on a trial and error
approach and little is known fundamentally about the relationships between the film
stress, hardness, yield strength, elastic modulus and the film thickness. The present work
contributes to the understanding of the sub-micrometer indentation tests that are now being
used to study the mechanical properties of these important thin films.

For over four decades, indentation hardness testing has been effectively used to deter-
mine the strength of materials very near the surface of the material. Hardness measurements
have been attractive because they can be made easily on a variety of materials. But, because
of the comparatively large indentations involved, even traditional microhardness tests are
not well suited for determining the properties of thin films with typical thicknesses in the
range of a few hundred angstroms to just a few micrometers. At present, very low load
microhardness testers[1–3], which produce indentation depths as small as a few hundred
angstroms, represent the most successful technique for studying the mechanical properties
of thin films. Although various fundamental mechanical properties appear to be obtainable
from these tests, very little effort has been made to understand the mechanics of the plastic
deformation involved in these sub-micrometer indentations. The lack of knowledge in this
area makes it difficult to determine the relative merits of various modifications that are
necessary for designing a typical thin film system. It is, therefore, important to conduct
continuum mechanics analyses in an effort to understand the underlying mechanics of these
indentation tests.

A few analytical treatments related to indentation mechanics exist in the literature.
The solution for a flat ended circular punch contacting a half space is well known[4]. Elastic
normal contact problems for layered media were analyzed by Chen and Engel[5] and
analyses of some punch problems for thin films on substrates were also given by King[6].
All of these solutions are applicable to elastic deformation only and cannot be used to

predict the complex elastic-plastic deformation involved in a typical indentation test. A few investigators[7-10] have used the finite element technique to analyze large indentations typical of the Brinell hardness test. The same method has also been used to calculate stresses in contact situations[11]. Semi-empirical analyses for indentation problems have been provided by other investigators[12-14]. Except for the work of Bourcier et al.[7], no detailed analyses have been made for extremely small indentations, such as those produced in typical sub-micrometer hardness tests[1, 3]. Due to the obvious mathematical complexities involved in such analyses, the finite element method is required and has been used in this investigation. The analysis involves a simulation of indentation tests, from which insight into the various mechanical properties of thin films can be gained. This work represents the first part of a more extensive theoretical research program aimed at understanding the mechanical properties of thin films on substrates. In the present study, the goals have been: (a) to demonstrate the feasibility of modeling a typical hardness test using the finite element technique and (b) to show how relevant basic mechanical properties can be extracted from such simulated results.

THE FINITE ELEMENT MODEL

Sub-micrometer indentation testing permits the measurement of force-distance relations on a very small scale; a detailed description of this can be found elsewhere[1, 3]. Simulations of these force-distance relations for the indentation of nickel, silicon and aluminum using a rigid indenter were performed using the large strain elasto-plastic feature of the ABAQUS finite element code[17], with uniaxial stress-strain data as input. The quasi-nature of the process permits us to use the static analysis performed by the program. Underlying the approach in this code is the discretization of the continuum involved: the indenter was considered to be perfectly rigid. Also, an important feature of this program involves the capability to model contact between the indenter and the sample as a sliding interface. The initial nodal gaps between the indenter and the surface of the specimen were prescribed and the program automatically keeps track of their change and indicates any gap closure or opening in a particular direction. These interface elements thus simulate contact between the indenter and the specimen surface. Whenever the closure distance between the indenter and the specimen becomes zero, contact is assumed and an external reaction force is exerted on this material point to keep it moving along with the indenter. Because the program calls for incremental loading and also makes use of interface elements, the expanding contact area associated with indentation occurs naturally whenever new interface elements come into contact.

In this analysis, the indenter and specimen are treated as bodies of revolution to avoid the inherent three-dimensional nature of the problem of indentation with a pyramid shaped indenter. If a pyramid indenter was used, it would have an elastic singularity at its edges but this would affect only the stress-strain response of a few of the elements adjacent to these edges. Thus, it is not expected to cause any noticeable difference in the load-deflection response of the material. Although this approximation is considered to be acceptable for the case of continuum plasticity, a three-dimensional analysis would be needed to treat crystal plasticity. Thus, in the present treatment the pyramid indenter was approximated by an axisymmetric cone of equal volume for a given indenter depth. The indenter and the specimen are shown schematically in Fig. 1, along with the appropriate boundary conditions for the problem. Symmetry properties have been used to simplify the boundary conditions. During preliminary simulations the boundary condition on the surface on the right-hand side of the specimen was changed from fixed radial displacements to traction free; this change had no effect on the indentation parameters, thus showing that this boundary was indeed remote. Because very small indentations were being simulated, the meshes near the indenter needed to be very fine to be able to describe the deformation and stress gradients associated with indentation with sufficient accuracy. Thus, extremely fine mesh sizes were used under the indenter; they became progressively coarser at distances further away from the indenter. Axisymmetric four node elements were used for the continuum. In order to obtain an accurate estimate of the radius of the contact area, an extremely fine mesh
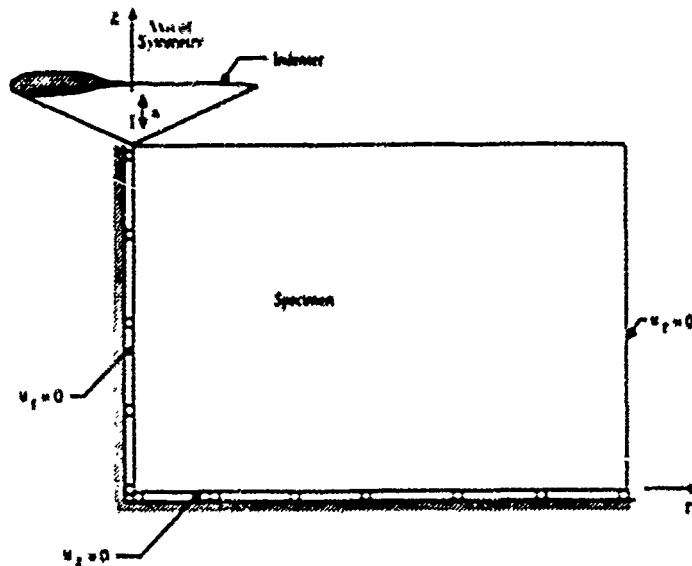
Fig. 1 Schematics of the specimen under the indenter showing the boundary conditions.

thickness of the order of 0.02 μm had to be used along most of the indenter contact surface. To keep the computer time within limits, a total of 461 elements including the interface elements were used for representing the deformed material. Figure 2 represents a magnified view of the elements near the indenter and the staircase arrangement for the other elements at points farther away from the indenter. When two elements are connected to a single element in this distribution, the middle node on the common face is constrained to lie on a straight line defined by the two corresponding end nodes.

To simulate a typical indentation process, a downward displacement (negative z-direction in Fig. 1) was imposed on the indenter; this causes the indenter to push into the surface of the material. Subsequently, the indenter was given an upward displacement until
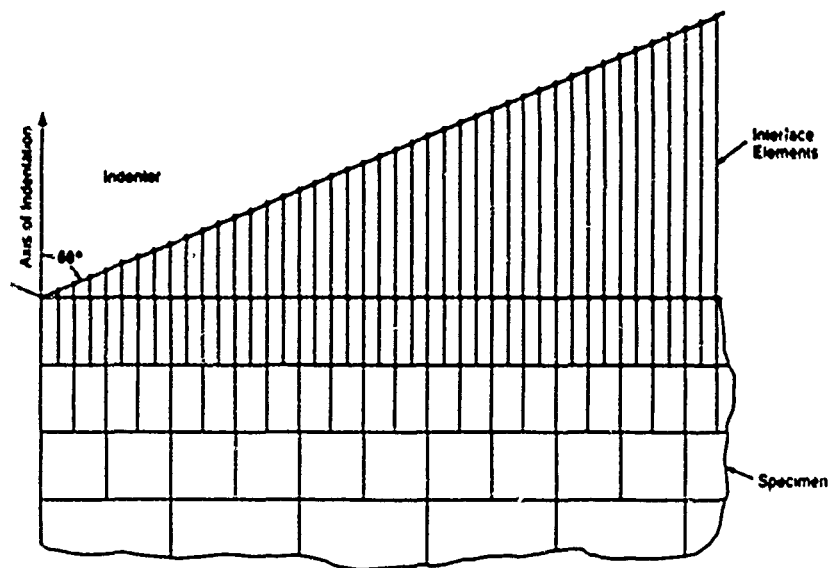

Fig. 2. Detailed pattern of mesh distribution near the indenter showing interface elements.

Table 1. Elastic and plastic properties of nickel, silicon and aluminum used in the analysis

Material	Young's modulus (GPa)	Poisson's ratio	Yield stress (MPa)	Strain hardening rate (MPa)
Nickel	207	0.31	350	300
Silicon	127	0.278	4410	0
Aluminum	75.9	0.33	485	144

it was free of contact with the specimen. For a given indenter displacement, the corresponding load determination was achieved by summing the reaction forces at the contact node points on the indenter. The interface between the specimen and the indenter was assumed to be frictionless since no noticeable change in the load–displacement response was observed by using a friction coefficient of 1. The mesh thickness of 0.02 μm along the indenter contact surface was determined by finding the mesh size below which no further significant changes in the indentation load–displacement response were observed. The hardness was also observed to remain constant with increasing indentation depth when the mesh was refined to this point (Figs 8 and 9).

The constitutive model for the specimen material (nickel, silicon and aluminum) was that of an elastic–plastic von Mises material with isotropic hardening. Two separate cases of strain hardening were considered, one with no strain hardening (i.e. the material was assumed to be elastic–fully plastic) and the other with a linear strain hardening rate of μ/200 for the material under consideration. The material properties used in the calculations are given in Table 1. The finite element calculations were performed using an IBM 4341 mainframe computer with run times of 15–25 h for average indentation depths.

RESULTS AND DISCUSSION

Results of the load–displacement simulation are shown in Figs 3 and 4. In these plots comparison is made between the experimental data obtained by Pethica et al.[1] for pure nickel and silicon and the simulated response from finite element calculations. The material properties for these simulations are given in Table 1. The agreement between our finite element analysis and experimental results is satisfactory, thus indicating the feasibility of using finite element analysis to describe the sub-micrometer indentation process. Slight numerical oscillations present in the load–depth curves are caused by discontinuous contact of the test specimen nodal points with the indenter surface. The differences between the
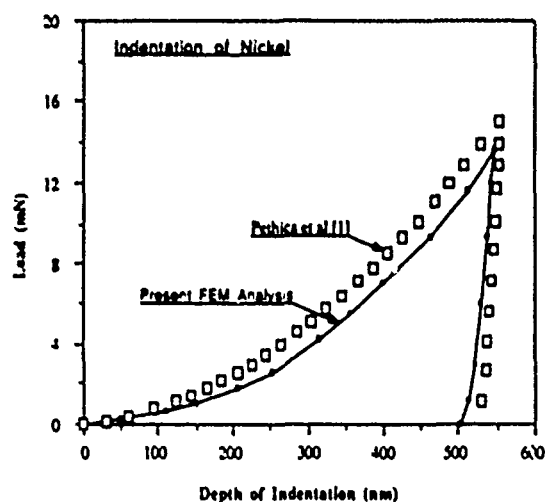

Fig. 3. Comparison between the results from the present FEM analysis and those from Pethica et al.[1] on indentation of nickel.
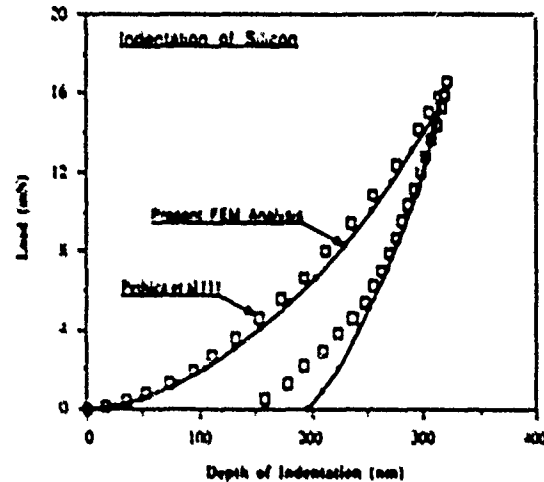
Fig. 4. Comparison between the results from the present FEM analysis and those from Pethica et al.[1] on indentation of silicon.

numerical and the experimental results may be due to differences in the actual and assumed yield stresses and work hardening rates of the materials involved. Also, the differences could be caused by differences between the tip geometry of the experimentally used indenter and the perfectly sharp indenter used in our calculations. The effect of tip geometry on the material response has been adequately discussed in Refs [1,3].

Having shown that the finite element analysis can adequately describe the experimental load-displacement response, we next compute the hardness and Young's modulus from these simulations and compare the results with the material data used as input for the calculations. We have also analyzed two completely different types of materials; a low yield strength, high strain hardening material represented by aluminum and a very high yield strength, low strain hardening material represented by silicon. Elastic moduli for the two materials were also very different. Simulated results for these materials are given in Figs 5-7. Strain hardening was not included in the silicon simulation (Fig. 5). For indentation of aluminium, both cases of no strain hardening (Fig. 6) and full strain hardening (Fig. 7) have been considered. Prescribing linear strain hardening for silicon did not produce any
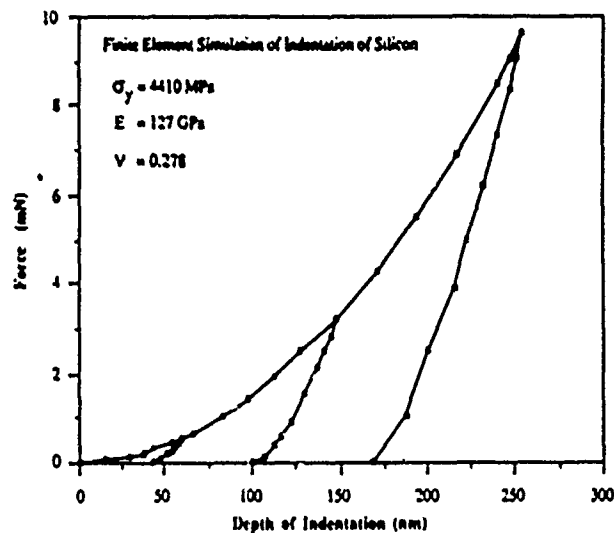

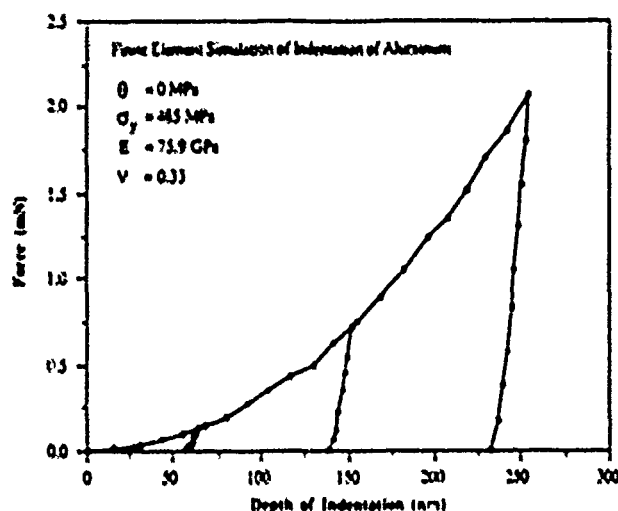Fig. 5. Load-depth response for indentation of silicon.

Fig. 6. Load-depth response of indentation of aluminum without strain hardening.

significant changes in the load-depth response. Unloading from three different points along the loading curve has been performed as shown. The general unloading response is similar to the experimentally observed response[1, 3]. As expected, silicon, with a high yield strength to elastic modulus ratio, shows a large amount of elastic recovery upon unloading whereas aluminum, with a low yield strength to elastic modulus ratio, shows little elastic recovery.

Hardness results

The microhardness is calculated as the load divided by projected area under the indenter at various points on the loading curve. These hardnesses are plotted as a function of indentation depth in Figs 8 and 9. The results show that the hardness is essentially independent of indentation depth. At smaller indentation depths, it was necessary to use a refined mesh size to obtain hardnesses that are independent of the depth of indentation over the range of depth considered. These results are expected for a homogeneous material described by a continuum based constitutive model. The calculations of hardness were performed using the load-depth results shown in Figs 5-7 and taking the value of contact areas directly from the FEM calculations.
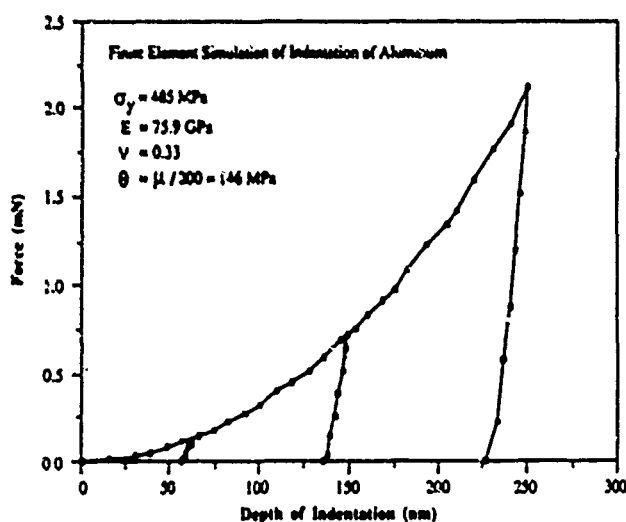

Fig. 7. Load-depth response for indentation of aluminum with strain hardening.
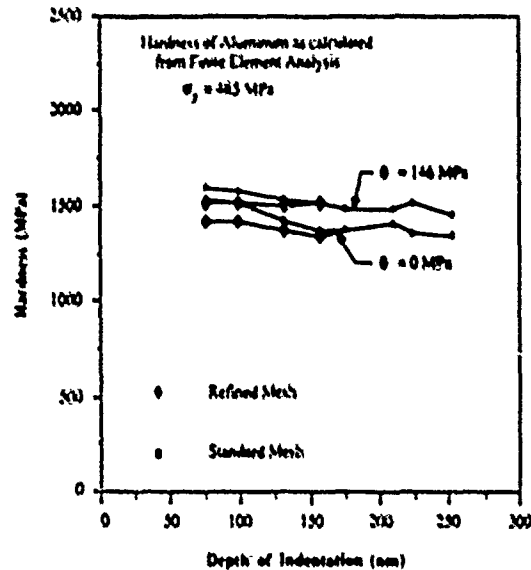
Fig. 8. Hardness variation as a function of depth of indentation for aluminum with and without strain hardening.

Determination of Young's modulus

As reported in Ref. [3], the slope of the unloading curve can be used as a measure of the elastic properties of the material being indented. If the contact area between the indenter and sample remains constant during initial unloading, the elastic behavior may be modelled as that of a flat ended cylindrical punch indenting an elastic solid. Loubet et al.[15] adopted the elastic solution given by Sneddon[16] and equated the projected area of contact with the indenter to the area of the punch. Obtaining an estimate of the true projected contact area under the indenter has been a source of some confusion because of the involvement of both elastic and plastic displacements under the indenter. The generally accepted procedure has been to estimate this contact area from the plastic depth of indentation. But, due to the elastic recovery of the indented material after unloading, the measured depth does not necessarily correspond to the plastic depth. It has been suggested[1] that for a
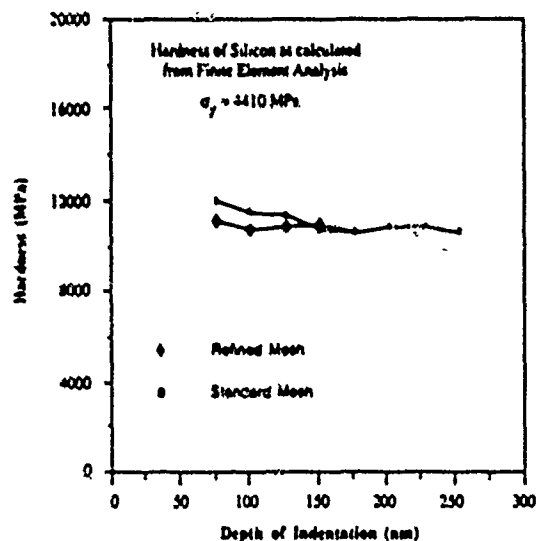

Fig. 9. Hardness variation as a function of depth of indentation for silicon.

material that exhibits a large amount of elastic recovery, the best estimate of the true, relaxed indentation depth is the depth at zero load, just before separation between the indenter and the material occurs. For a material that exhibits only a small amount of elastic recovery, the plastic depth can be taken as the depth at the highest load point, before unloading begins. An alternate suggestion in Ref. [3] is that the plastic depth for all materials is that point at which an extension of the linear unloading curve intersects the depth axis. These various possibilities for estimating the plastic depth and the corresponding contact area from the load–depth response are analyzed below.

Following the same analysis as proposed by Loubet et al.[15], but using the plastic depth, h_p, instead of the diagonal length used by them and assuming a perfectly rigid indenter with an ideal pyramidal geometry (having the same depth–area relationship as the Vicker's indenter), Ref. [3] obtained the following relationship for the elastic modulus:

$$E = \frac{(1 - v^2)}{2h_p} \sqrt{\left(\frac{\pi}{24.5}\right)} \frac{dP}{dh} \tag{1}$$

where h_p is the plastic depth, dP/dh the slope of the unloading curve and E and v are Young's modulus and Poisson's ratio, respectively. Below we use eqn (1) to estimate Young's modulus from the simulated load–depth curves.

The finite element results showed that for the first three points (on average) in the unloading curve, the indenter and the sample remained in contact. This suggests that it is valid to use eqn (1) to determine the elastic modulus of the specimen from the unloading slope. Because the contact area can be calculated directly from the FEM results, it is not necessary to use the plastic depth in determining the elastic modulus from the unloading slope. Rewriting eqn (1) in terms of the projected contact area, A, with the indenter, we obtain

$$E = \frac{(1 - v^2)}{2} \sqrt{\left(\frac{\pi}{A}\right)} \frac{dP}{dh} \tag{2}$$

as an alternate expression for the elastic modulus.

With eqns (1) and (2) one can obtain the elastic modulus of the material being indented from the unloading slope dP/dh using either the plastic depth, h_p, or the projected contact area, A, with the indenter. As discussed above, the plastic depth may be determined from the intercept of the linear portion of the unloading curve on the depth axis or it may be taken to be either the depth at maximum load or the final depth in the unloaded state. As noted before, the projected contact area is found by knowing the radius of contact from the FEM calculations. To obtain a good estimate of dP/dh from the simulated load–displacement curves (Figs 5–7), a straight line was fit through the first three data points on the unloading curves. The highest load point at which the unloading begins was included in this fitting procedure. More points on the unloading curve with full contact between indenter and material could be obtained by using very small increments in displacement but at the expense of a large amount of computer time.

Table 2 summarizes various calculations of Young's modulus for aluminum and silicon using the methods described above. Out of the four calculated Young's modulus values, the first column involves the use of the flat punch model recast in terms of projected area of contact (eqn (2)) and the others involve the use of eqn (1). The second column makes use of the intercept of Ref. [3] to obtain the plastic depth. The third column is based on the assumption that the plastic depth is the final depth, i.e. the depth of indentation where the unloading curve reaches zero load value as proposed by Pethica et al.[1]. The fourth column involves the assumption that the plastic depth is the full depth of indentation at the point of maximum load. The purpose of these different types of calculations was to determine the effect of these various measures of plastic depth on the determination of Young's modulus. As seen from these calculations, the values of Young's modulus obtained using

Table 2 Comparison between theoretical and calculated values of Young's modulus

Depth of indentation (nm)	Flat punch model	Calculated Young's modulus (GPa)		
		Extrapolated depth model	Final depth model	Full depth model
Material: silicon with no strain hardening (E = 127 GPa)				
254.0	131.0	121.0	143.2	96.4
152.4	123.6	121.4	137.2	89.1
61.8	126.9	120.8	137.3	88.9
Material: aluminum with no strain hardening (E = 75.9 GPa)				
242.06	77.2	72.2	74.0	68.3
151.05	76.9	71.0	72.4	66.1
61.65	76.8	71.6	73.2	65.4
Material: aluminum with full strain hardening (E = 75.9 GPa)				
252.18	77.5	72.9	73.9	67.3
149.52	77.3	71.7	74.7	67.4
61.65	76.5	71.4	72.0	65.4

the flat punch model coupled with the projected contact area from the FEM calculations (eqn (2)) are in very good agreement with the known values of Young's modulus for both materials. None of the methods based on plastic depth (eqn (1)) give equally good results. The extrapolated plastic depth model gives values that are near to but less than the known values of elastic modulus for both materials. Values of Young's modulus based on the final depth model are higher than the known value for silicon and less than the known value for aluminum. The full depth model gives values that are much less than the known values for both materials. Considering the results for both materials, the extrapolated depth model appears to give the best estimate of the projected contact area and the best procedure for determining the elastic modulus from loading and unloading curves alone.

Effect of elastic and plastic properties on hardness

So far we have not discussed the separate effects of elastic and plastic properties on the hardness of a material. Normally, the mean contact pressure under the indenter, p_m, is related to the yield stress of the material under compression, σ_y, by a generalized expression based on the deformation theory of a rigid-perfectly plastic solid[12] as

$$p_m = C\sigma_y \tag{3}$$

where C is a constant having a typical value close to 3. For ductile metals and other similar materials, this relationship works well. But, indentation experiments with highly elastic materials, such as polymers, have shown that the elastic and plastic strains associated with indentations are of the same order of magnitude and the above relationship does not apply. Johnson[13], using Marsh's[18] model of the expanding spherical cavity in an elastic-perfectly plastic solid derived a comprehensive relationship between the hardness, yield stress, Young's modulus and the indenter shape defined by the angle β between the indenter and the horizontal. For the case of an incompressible material, his results simplify to

$$\frac{H}{\sigma_y} = \frac{2}{3}\left[1 + \ln\left(\frac{1}{3}\frac{E}{\sigma_y}\tan\beta\right)\right]. \tag{4}$$

He compared this prediction with experimental data for various kinds of materials and showed that for an indenter with a given geometry, the hardness of a material for an elastic-plastic material is not only a function of yield stress but also a function of the parameter E/σ_y, where E is the Young's modulus of the material.
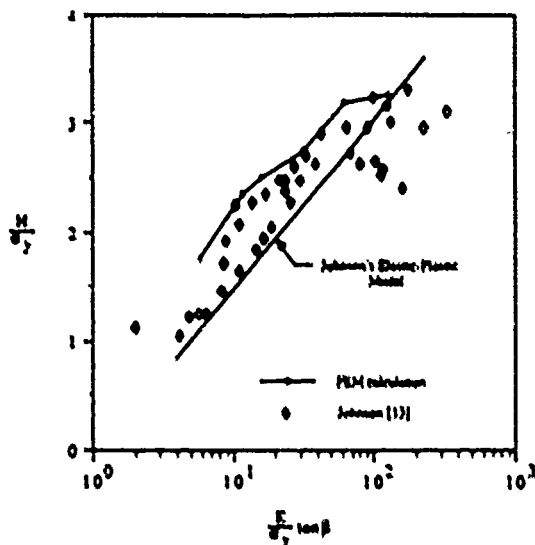
Fig. 10. Comparison of various experimental hardness results[13] with the predictions of Johnson's elastic-plastic model and the present FEM analysis.

In Fig. 10, we have compared the experimental data compiled by Johnson[13] with a few values of the hardness obtained from the present finite element analysis. Although the Johnson model correctly predicts the general trend in the data, it does not correlate the data well at larger values of $E\sigma$, where the data seem to reach a plateau, with H/σ values of about 3. This corresponds to the value predicted by rigid plastic theory from classical mechanics. The FEM analysis predicts an upper bound to the experimental results and also describes the plateau behavior at large values of E/σ_y. A more detailed analysis of the various predictions of the Johnson model is being made[19] and will be described elsewhere.

CONCLUSIONS

The results obtained from this study indicate that it is possible to successfully simulate the overall load-depth response of a sub-micrometer indentation test for different types of materials by using the finite element technique together with simple constitutive data as program inputs. It has also been demonstrated that one can obtain Young's modulus for the material from the slope of the linear portion of the unloading curve in such a simulation. This has been shown earlier in Ref. [3] and applied to actual test data. In addition, the FEM analysis yielded a hardness curve which was more or less independent of the depth of indentation. It is thus possible to obtain hardness values for the material from simulated data along the loading curve. The present study also gives theoretical justification for the use of the extrapolated depth as the best measure of the plastic depth of indentation. Finally, it has been shown that the FEM analysis predicts the response for an elastic-plastic material with respect to the relationship between hardness, yield stress and the Young's modulus of the material.

REFERENCES

1. J. B. Pethica, R. Hutchings and W. C. Oliver, Hardness measurement at penetration depths as small as 20 nm. Phil. Mag. (A) 48, 593 (1983).

2. N. Gane and J. M. Cox. The micro-hardness of metals at very low load. *Phil. Mag.* **22**, 881 (1970).
3. M. F. Doerner and W. D. Nix. A method of interpreting the data from depth-sensing indentation measurements. *J. Mater. Res.* **4**, 601 (1986).
4. S. Timoshenko and J. Goodier. *Theory of Elasticity*. McGraw-Hill, New York (1970).
5. W. Chen and P. Engel. Impact and contact stress analysis in multilayer media. *Int. J. Solids Structures* **8**, 1257 (1972).
6. R. B. King. Elastic analysis of some punch problems for a layered medium. *Int. J. Solids Structures* **23**, 1657 (1987).
7. R. J. Bourcier, C. M. Stone and F. O. Yost. A numerical and experimental study of the indentation hardness test. Sandia Report No. SAND 85-0486 (1985).
8. C. H. Lee, S. Masaki and S. Kobayashi. Analysis of ball indentation. *Int. J. Mech. Sci.* **14**, 417 (1972).
9. P. S. Follansbee and G. B. Sinclair. Quasi-static normal indentation of an elasto-plastic half-space by a rigid sphere—I. *Int. J. Solids Structures* **20**, 81 (1984).
10. C. Hardy, C. N. Baronet and G. V. Tordion. The elasto-plastic indentation of a half-space by a rigid sphere. *Int. J. Numer. Meth. Engng* **3**, 451 (1971).
11. A. G. Tangena and P. A. M. Hurkx. Calculations of mechanical stresses in electrical contact situations. *IEEE Transactions, Components, Hybrids and Manufacturing Technology*, Vol. CHMT-8, No. 1, 13-20, March (1985).
12. D. Tabor. *The Hardness of Metals*. Clarendon Press, Oxford (1951).
13. K. L. Johnson. The correlation of indentation experiments. *J. Mech. Phys. Solids* **18**, 115 (1970).
14. C. Rubenstein. A critical appraisal of static hardness measurements. *J. Appl. Mech.* **48**, 796 (1981).
15. J. L. Loubet, J. M. Georges, J. M. Marchesini and G. Meille. Vicker's indentation curves of magnesium oxide (MgO). *J. Tribology* **106**, 43 (1984).
16. I. N. Sneddon. The relation between load and penetration in the axisymmetric Boussinesq problem for a punch of arbitrary profile. *Int. J. Engng Sci.* **3**, 47 (1965).
17. ABAQUS finite element program. HKS Inc., Providence, Rhode Island (1985).
18. D. M. Marsh. Plastic flow in glass. *Proc. R. Soc.* A279, 420 (1964).
19. A. K. Bhattacharya and W. D. Nix. Unpublished research.

Mechanical deflection of cantilever microbeams: A new technique for testing the mechanical properties of thin films

T. P. Weihs, S. Hong, J. C. Bravman, and W. D. Nix
Department of Materials Science and Engineering, Stanford University, Stanford, California 94305

The mechanical deflection of cantilever microbeams is presented as a new technique for testing the mechanical properties of thin films. Single-layer microbeams of Au and SiO_2 have been fabricated using conventional silicon micromachining techniques. Typical thickness, width, and length dimensions of the beams are 1.0, 20, and 30 μm, respectively. The beams are mechanically deflected by a Nanoindenter, a submicron indentation instrument that continuously monitors load and deflection. Using simple beam theory and the load-deflection data, the Young's moduli and the yield strengths of thin-film materials that comprise the beams are determined. The measured mechanical properties are compared to those obtained by indenting similar thin films supported by their substrate.

I. INTRODUCTION

With the advent of modern integrated circuitry a new field of research concerning the mechanical behavior of materials has developed. The impetus for this development has been the realization that the manufacture and performance of very large scale integration (VLSI) devices can be influenced by the deformation of materials used in their fabrication. Under various conditions metal interconnects are found to peel and deform while dielectric coatings have been observed to crack and buckle. Such phenomena, which may adversely affect the performance of circuits, cannot be predicted easily using the properties of bulk materials. The small and continuously decreasing dimensions of features in integrated circuits forces deformation to occur on a scale not yet studied. This constraint can alter the properties of interconnect and dielectric materials compared to those that were measured on bulk samples. Consequently it has been necessary to develop new experimental techniques to characterize and investigate the mechanical properties of materials on a micron and submicron scale. Submicron indentation testing, wafer curvature testing, bulge testing, microtensile testing, and resonant frequency testing have proven effective in this endeavor. However, all of these techniques have inherent limitations that restrict their utility in experimental studies.

The submicron indentation of thin films on substrates can lead to quick and accurate measures of hardness, but the large pressures under the indenter may alter the structure of the thin film being tested.[1,2] Determining the elastic moduli of materials from indentation data is possible but not simple for hard films, and

the influence of the substrate must be considered. With wafer curvature techniques the average stress and strain in a film can be measured, but the range of stresses is limited by the thermal expansion and/or growth mismatch of the substrate and the film.[3-5] Thus, for a given film and substrate, one cannot dictate the stress to be applied to the film. Furthermore, the measured stress is an average value for a large part of the wafer and this may mask significant, local fluctuations. The bulge test typically uses a smaller sample than the wafer curvature technique, but it too fails to provide localized information.[6,7] Another common drawback to the bulge test is the need to handle samples prior to testing. With thin films, plastic bending or stretching of a sample is difficult to avoid while removing it from the substrate and mounting it in the test apparatus. The same problem is encountered during microtensile testing.[8] Lastly, while the resonant frequency measurements of elastic moduli of cantilever beams avoid the need to handle samples and can be local, they are limited to elastic measurements and the range of experimental error can be large.[9]

In an effort to avoid some of these difficulties, we have developed a new experimental technique based on the deflection of free-standing cantilever microbeams. Briefly, the technique involves the fabrication of the beams using silicon micromachining techniques and the deflection of the beams using a Nanoindenter, a load-controlled submicron indentation instrument (made by Nano Instruments, Inc. and marketed by Microscience, Inc.). The Nanoindenter mechanically bends the beams while continuously monitoring the loading and the deflection of the beams. There are several advantages inherent in this technique. First, all effects of the substrate

are removed since only the thin films comprise the cantilever beams. Second, the beams themselves are never directly handled. Third, the measurements are local in scale since the beam dimensions are typically on the order of microns. Fourth, both elastic and plastic properties—namely Young's modulus and yield strength—can be studied, and, fifth, experimental error is reasonably low. Lastly, a large assortment of thin film materials can be prepared for testing using conventional fabrication processes.

II. SAMPLE PREPARATION

Three materials that are frequently used in microelectronics, both in chip and packaging levels, were selected for this study. Thermal SiO_2 and low temperature oxide (LTO) were chosen as representative dielectric and passivation materials used in integrated circuits, and gold was chosen as a representative interconnecting material used in packaging. To fabricate test structures of thermal SiO_2, LTO, and gold, thin films of these materials were grown or deposited on lightly doped p-type (100) silicon wafers.

Wet-thermal SiO_2 layers with thicknesses of 1.10 and 1.51 μm were grown at 1100 and 1000 °C, respectively. The thicknesses of the oxides films were measured with an ellipsometer. The films were then patterned using standard photolithography processes and a wet etchant of buffered hydrofluoric acid (HF). After patterning the films, the exposed Si was etched in an anisotropic etchant, rinsed, and then dried. Prior to the patterning of the thin films and the etching of the Si, the straight side of the mask pattern was aligned parallel to the (110) direction of the substrate to minimize undercutting of the final test structures at concave corners. The final test structures consist of cantilever beams extending over an open trench that is bounded by four {111} sidewalls (see Fig. 1). The anisotropic etchant used for the Si was an EDP (66 wt.% ethylendiamine, 23 wt.% H_2O, 11 wt.% pyrocatechol) solution heated to 115 °C. This particular etchant was utilized because it maximizes the difference in etching rates between the (100) silicon and SiO_2. The etching rate of Si in the (100) direction is about 1 μm/min, while that of thermal SiO_2 is 2 Å/min. With a 60 min etch in EDP it is routinely possible to fabricate beams 20 μm in width and up to 150 μm in length. Figure 1 is a scanning electron microscope (SEM) photograph of a test pattern of thermal SiO_2 cantilever beams with various lengths. The widths and thicknesses of the beams are 18 and 1.10 μm, respectively.

The LTO films with thicknesses of 1.35 μm were deposited by a commercial low-pressure chemical vapor deposition (LPCVD) system at 450 °C and 280 mTorr in a mixture of SiH_4 and O_2. The films were subsequently annealed at 950 °C in an O_2 atmosphere for 30 min to densify the as-deposited structure and to relieve the internal stresses in the film. By repeating the procedures described above, the final test structures were fabricated. The etching rate of LTO in EDP is similar to that of SiO_2.

For the patterning of Au films, a "lift-off technique" was utilized. Photoresist was applied, exposed to UV light, and then soaked in chlorobenzene for 5 min to form hangovers in the sidewall profile of the photoresist. Gold was then deposited on this prepatterned photoresist by employing an e-beam evaporation system. Warm acetone was used to dissolve the photoresist and to lift off the unwanted Au film: Hence, no wet etching of the Au was needed. To improve the adhesion of the Au films, a very thin (50 Å) layer of Cr was deposited prior to the Au. The metals were deposited at room temperature in a vacuum of 2×10^{-6} Torr and at a rate of about 50 Å/min. The thickness of the Au film was measured to be 0.87 μm using a surface profiler. Finally the exposed Si substrate was anisotropically etched. Although it was too difficult to measure, the etching rate of Au in EDP is thought to be comparable to that of thermal SiO_2 and LTO.

The Au test pattern is identical to that used for the oxide beams. However, residual stresses with nonuniform distributions developed in the Au films during their growth. As a result, when the substrate was etched away, the beams bent downward slightly, as seen in Fig. 2 (an analysis of residual stresses in the free-standing beams is presented in Appendix A). One should note
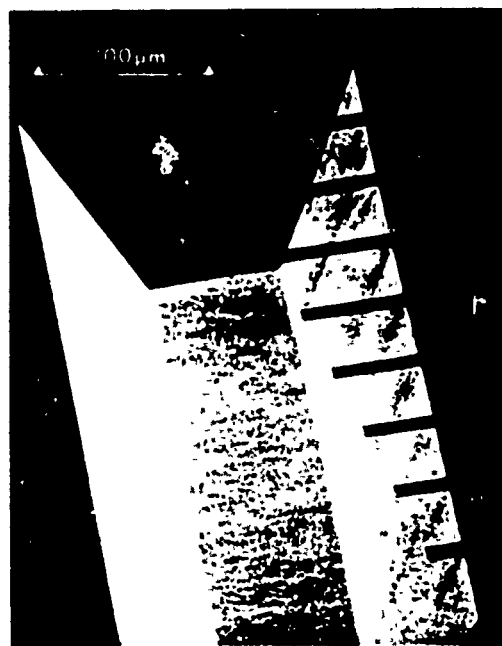

FIG. 1. Scanning electron micrograph of thermal SiO_2 cantilever microbeams extending over a Si etch pit.
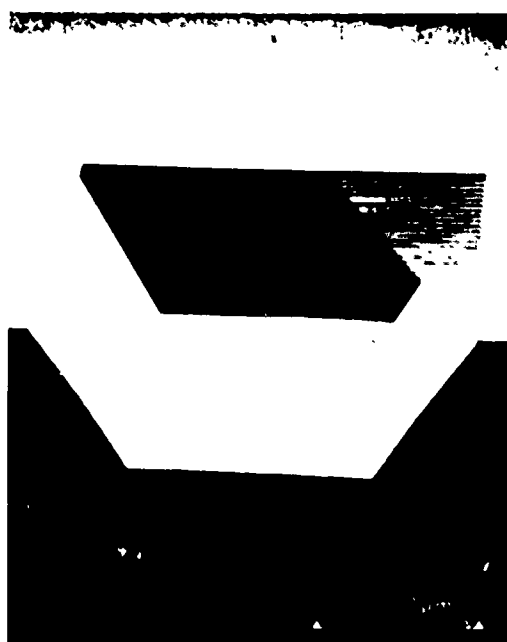
FIG. 2. Scanning electron micrograph of Au cantilever microbeams extending over a Si etch pit

FIG. 3 Schematic of a Nanoindenter loading mechanism and a cantilever microbeam

that this bending is not due to the weight of the beams. This conclusion is obtained using simple beam theory and the density of the test material.

III. TEST PROCEDURES

After fabrication the wafers were cut into small (1.5 × 1.5 cm) sections and were mounted on Al blocks (1.5×2.0×3.0 cm) with crystal bond wax. A mounted sample was then positioned under the Nanoindenter. The Nanoindenter itself is load controlled and depth sensing with a 0.25 μN load resolution and a 2–3 Å depth resolution. Information regarding its components and operation have been given elsewhere.[1,2] We mention here three facts concerning the operation of the Nanoindenter. First, as reported earlier[1,2] the Nanoindenter was originally developed to indent thin sections of material. However, in the technique described here, the Nanoindenter is used to mechanically deflect microbeams The indentation of the beams during their deflection is a minor effect that cannot be avoided. Its significance will be discussed later. The second fact is the 0.5 μm translational resolution of the stepping motors connected to the sample mounting. These motors allow accurate horizontal (x and y) translations of the sample so that a microbeam 10–20 μm wide can be point loaded at various positions along its length. A schematic diagram of such a loading is shown in Fig. 3. Finally, one should also note that supporting springs are attached to the shaft of the indenter in Fig. 3. Consequently, when the
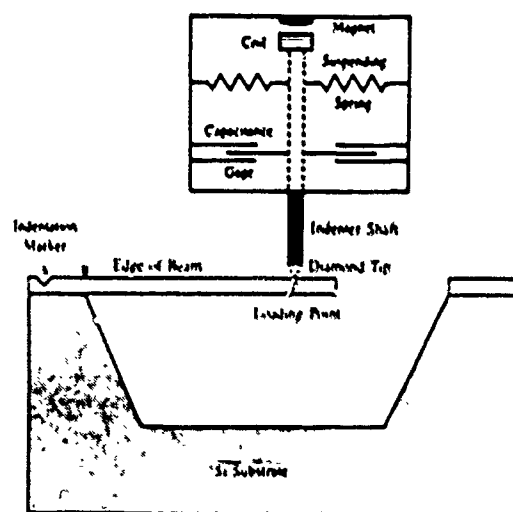
diamond tip displaces a microbeam, the loading mechanism is actually displacing two springs in parallel—the microbeam and the spring suspending the shaft of the Nanoindenter.

During testing, the microbeams were deflected at a constant velocity between either 3 and 6 nm/s or between 30 and 60 nm/s. The higher deflection rates were employed so that the loading rate on the beam would match the loading rate typically used to indent films on substrates. After a predetermined deflection was reached, the load was held constant for 10 s and then the beam was unloaded at rates similar to those for loading. All tests were conducted at room temperature.

In order to accurately position the beams for testing, loading points along the beams (point C in Fig. 3) and the fixed end of the cantilever beams (point B) were prerecorded using the x and y stepping motors. In addition, the position of an indentation in the film (point A) was prerecorded. This indentation allowed the actual deflection length c (to be called the effective length) to be checked after testing. By measuring the distance between the indentation and the fixed end of the beams, and by comparing the measured value with the prerecorded distance (AB), changes in the effective length due to the thermal expansion of the xy table and sample mounting could be determined. In addition to the above checks of the effective loading lengths, small indentations marking the loading points on heavily loaded Au beams were seen in a scanning electron microscope (SEM). These indentations allowed an additional measurement of the effective lengths of the Au beams.

To compare the microbeam technique to a thin-film mechanical testing technique already reported, the films that comprise the microbeams were indented in areas

supported by substrate. Using the Nanoindenter the films were indented at room temperature and at a constant velocity between 3 and 6 nm/s. The indentations themselves varied in total depth between 20 and 400 nm or between 2% and 50% of the film thickness. By applying the technique developed by Doerner and Nix[1] to the indentation data, the Young's moduli of the films and the hardnesses as a function of depth were determined.

IV. THEORY

The theory of the elastic deflection, w, of a cantilever beam with a rectangular cross section has been well developed.[10] For small deflections of a thin beam,

$$w = 4Pc^3(1 - v^2)/(bt^3E), \tag{1}$$

where P is the force applied, c is the effective length, v is Poisson's ratio, b is the width, E is Young's modulus, and t is the thickness of the beam. Equation (1) indicates that the elastic deflection of a microbeam will vary linearly with the force. Furthermore, it suggests that Young's modulus of the beam can be determined from the slope of this linear relation. However, as mentioned in the previous section, the Nanoindenter records the load required to displace both its own spring and the microbeam. Consequently, the load carried by the springs of the Nanoindenter must be subtracted from the total applied load to determine the load on the beam. In order to do this accurately the spring constant of the microbeam must be approximately equal to or greater than 10% of the spring constant for the Nanoindenter. Otherwise, the additional load required to bend the beam (while deflecting the spring) will be lost in the noise associated with the total load voltage measurements. For a typical beam (20 μm width, 1 μm thickness) employed in this study, this constraint prevents the effective length from being more than twice the width of the beam. For effective lengths greater than this, the beams are typically too compliant to be resolved accurately during loading.

The application of simple beam theory to the load-deflection data of beams also enables one to determine the yield strength of the microbeam material. For a homogeneous, cantilever beam under load, the strain in the beam varies linearly through the thickness so that the maximum stress at a given length occurs at the top and at the bottom of the beam. For the loading shown in Fig. 3, the top of the beam is in tension and the bottom of the beam is in compression. In addition, the maximum stress in the beam is located at the fixed end of the beam where the applied moment is greatest (point B in Fig. 3). When this maximum stress reaches the yield strength of the material, the beam begins to deform plastically. The onset of such deformation can be recognized in the plot of force versus deflection by a deviation from

linearity. The load that marks this deviation is defined as the yield load, P_y, and the yield strength is given by

$$\sigma_y = 6cP_y/bt^2. \tag{2}$$

Following yielding, the beam continues to bend as more load is applied, but some of the strain in the beam is plastic and thus nonrecoverable. When the beam is fully unloaded, it does not return to its initial position of zero displacement.

The shape of the force versus deflection curves after yielding and prior to unloading depends on both the elastic and plastic properties of the material. For the case of an elastic–plastic material that shows little or no strain-hardening, the shape of a force versus deflection curve can be easily modeled. By relating the applied moment to the curvature of the beam, and by integrating the curvature equation to obtain the beam deflection, a force versus deflection relationship can be found.[10] Figure 4 shows a load-deflection curve obtained by such an analysis. According to this analysis, the maximum load that can be supported by the beam is 1.5 times the yield load.

For a beam composed of a material that exhibits strain-hardening, the maximum load that can be supported by the beam is higher than that for an elastic-plastic material. To model this kind of behavior we assume a parabolic strain-hardening law of the form

$$\sigma = \sigma_y + K\epsilon_{pl}^N, \tag{3}$$

where σ_y is the yield strength of the material, K is a material parameter, ϵ_{pl} is the plastic strain, and N is the strain-hardening exponent, taken to be 0.5 in the present analysis. To determine the shape of the force versus de-
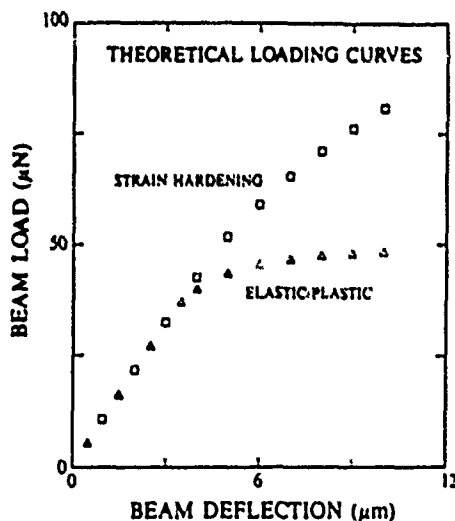

FIG. 4. Theoretical load-deflection curves for cantilever beam deflections. The lower curve represents an elastic–plastic material while the upper curve represents a material that strain hardens.

flection curve for such a material a different analysis is used. Instead of calculating deflections from applied moments, forces are calculated from deflections. More precisely, the total strain at the fixed end of the beam is determined for a given deflection using the following equation:

$$\epsilon_f = (3tw/2c^2)(2y/t).$$ (4)

To determine the variation of stress across the thickness of the beam, the total strain, which varies linearly across the thickness of the beam, is divided into the plastic and elastic components.

$$\epsilon_f = \epsilon_{el} + \epsilon_{pl}.$$ (5)

Then, for deflections beyond the yield point, equations relating stress to strain can be substituted into Eq. (5).

$$\epsilon_f = \sigma/E + K'(\sigma - \sigma_y)^{1/\gamma}.$$ (6)

Using Eq. (6) the normal stresses are determined for a given deflection and then integrated over thickness to obtain the applied force.

$$F = \left(\int y^\circ \sigma(b\,dy)\right)c^{-1}.$$ (7)

The result of such an analysis is presented as curve A in Fig. 4 (and it is also compared to experimental results in Fig. 10).

V. RESULTS

A. Elastic beam deflections

Figure 5 presents typical load-deflection data for the bending of an oxide beam. The initial linear section of the data represents only the loading of the Nanoin-

denter springs that support the indenter. After a displacement of approximately 0.5 μm, the diamond tip contacted the SiO_2 beam. At this point the load was required to both bend the beam and deflect the springs supporting the shaft of the indenter. Thus the slope of the curve increases noticeably. Curves labeled A and B that emanate from the same point are two different deflections of the same beam. As expected, loading A, which had a smaller effective length, produced a larger force-displacement slope than loading B. To calculate Young's modulus of this beam the spring constant for the indenter was first determined by performing a regression analysis on the data from the initial part of the curve. Once measured, the spring constant, equal to 59.4 N/m, was subtracted from the slopes of curves A and B. The net slopes were then used in conjunction with the measured dimensions of the sample to determine Young's modulus of the beam. Figure 6 shows the net loading and displacement of the SiO_2 beam.

In general the deflection of both LTO and thermal oxide beams was linear and elastic. The unloading curves consistently overlapped the loading curves with no signs of permanent deformation. Several beams were deflected to distances greater than 1/2 their effective lengths. For such tests the elastic strains at maximum load were as high as 0.01. Such straining without noticeable cracking suggests that the oxides have much higher strengths as thin films than they do as bulk materials. Using a Griffith analysis, the approximate maximum flaw size can be shown to be less than 140 nm.
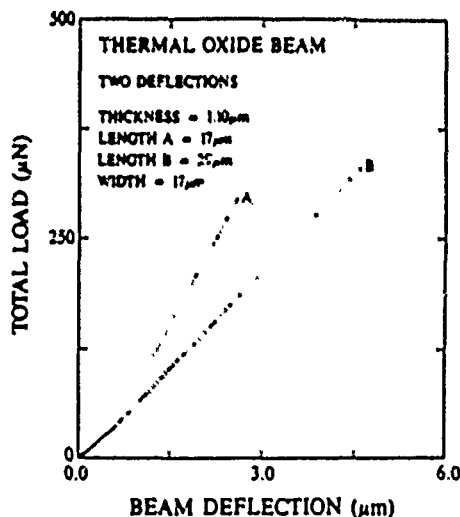

FIG. 5 Total applied load and total displacement measured for two deflections of one thermal SiO_2 cantilever microbeam
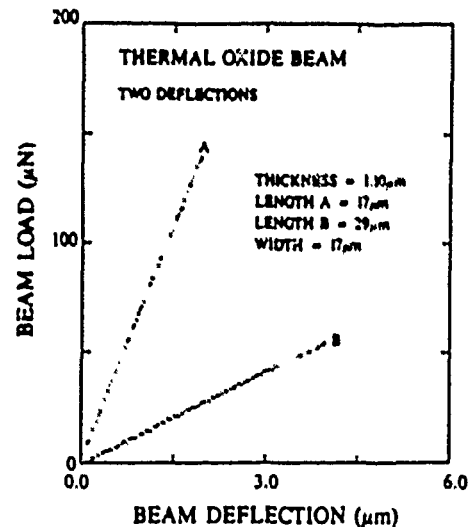

FIG. 6. Net loadings and net displacements of the thermal SiO_2 microbeam described in Fig. 5

B. Elastic-plastic beam deflections

Load-deflection curves for a Au microbeam with a 50 Å Cr under layer are shown in Figs. 7 and 8. Figure 7 shows the first bending of the microbeam and Fig. 8 shows the second bending of the beam at the same effective length. At the start of the first loading the load-displacement curve is linear, and deformation is thought to be elastic. As the deflection continues, the slope of the curve begins to decrease when the Au deforms plastically. Plastic deformation at the fixed end of the beam is verified by the permanent deflection of the beam observed after unloading. The permanent deflection after testing can be seen using an optical or scanning electron microscope.

When the same beam is deflected a second time in the same location, the loading curve is linear up to the maximum load that had been applied in the first test (Fig. 8). This also suggests that the Au beam yielded during the first test and thereby raised the beam's yield load. With further displacement in the second test, the curve became nonlinear again as the Au continued to deform plastically. The deflections of other Au beams showed similar elastic and plastic behavior. The variation in loading rate between 3–6 nm/s and 30–60 nm/s had no noticeable effect on the behavior or measured properties of the Au or oxide beams.

VI. DISCUSSION

A. Measurement of Young's modulus

Table I shows average values of Young's modulus obtained from beam deflections, from indentation tests,

and from literature.[9,11-13] In calculating the measured moduli, Poisson's ratios of 0.16 and 0.42 were assumed for the oxides[14] and Au,[15] respectively. The Young's modulus of thermal oxide measured by deflections of beams is 64 GPa ± 10% while the corresponding Young's modulus for LTO is 44 GPa ± 5% and for Au it is 57 GPa ± 11%. These experimental values agree with other thin film measurements. For the thermal oxide, agreement is found with experiments performed using a resonant frequency measurement technique[9] and a bulge testing technique.[11] The value of Young's modulus for Au agrees with microtensile tests.[12] However, the elastic properties obtained by the deflection of thin beams do not agree with values measured by indentation of the same thin films. To explain this disagreement, two possible explanations are considered.

The first explanation is densification of material under the the diamond tip during indentation experiments. Comparing the two techniques and their corresponding results, the densification of both the LTO and the thermal oxide under the diamond tip upon loading could stiffen the materials. Consequently, the unloading curves for the indentation experiments would produce higher values of Young's modulus than the beam deflections as seen. Comparing the two different oxides, the thermal SiO₂ is expected to be stiffer since it had more time (2.5 h) at a higher temperature (1100 °C) than the LTO, which was annealed at 950 °C for only 30 min. The results of beam deflections support this hypothesis. However, if the two materials were compressed to equal densities during indentation loading, then the unloading curves would produce similar Young's moduli for both
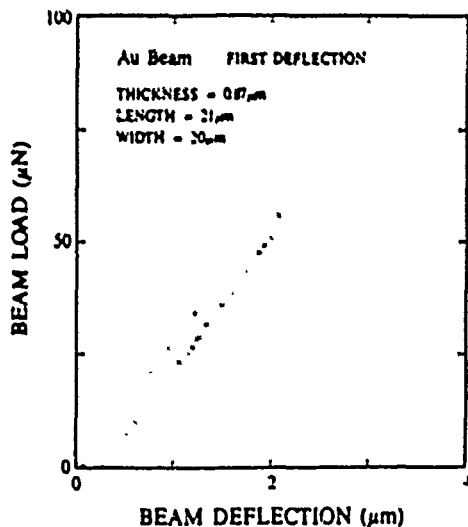
FIG. 7. Net load and net displacement measurements for the first deflection of a Au cantilever microbeam.
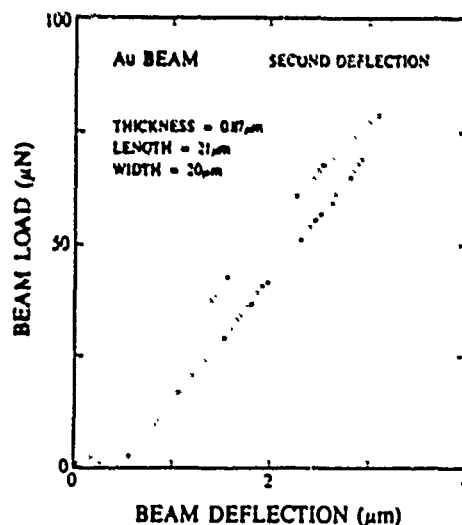
FIG. 8. Net load and net displacement measurements for the second deflection of a Au cantilever microbeam.

TABLE I. Young's modulus, yield strength, and hardness of thin Au and SiO₂ films.

| | Beam test | | Indentation test | | Published literature |
	E	σ₋	E	Hardness	E
Thermal SiO₂	64	>0.57	83	12	57(Ref. 9), 66(Ref. 11)
LTO	44	· ·	85	10	· · ·
Au	57	0.26	74	1	39-78(Ref. 12), 80(Ref. 13)

oxides as obtained. Thus both a comparison of the two techniques and a comparison of the two oxides support the argument that densification of the oxides during indentations may raise the Young's moduli measured by such a technique. In the case of the Au films, though, no densification during indentations is expected.

The second possible explanation for the differences in the values of Young's moduli obtained by the two techniques is changes in Young's moduli due to the large pressures under the indenter. For the oxide films, little or no increase in Young's modulus is expected, but for the Au thin films the 1.0 GPa pressure directly under the indenter could increase Young's modulus by up to 4.0%.[15] However, such an increase cannot explain the larger difference in the measured values in Table I. Consequently, no complete explanation is offered for the Au thin films at this time.

Figure 9 shows measured Young's moduli for thermal SiO₂ obtained from different tests on various microbeams. Since the average value of Young's modulus remains constant, even as the length to width ratio decreases from 2 to 1, the simple elastic beam theory invoked in this study appears to be appropriate. One can

also notice in Fig. 9 that the spread of data for the elastic modulus of the mal oxide increases as the ratio of width to length decreases. This spreading is attributed to the fixed error associated with the measurement of the geometry of the microbeams. In measuring the effective lengths of the beams, whether by SEM analysis or by checking prerecorded distances, we assume a 0.5 μm error. For a typical length employed in our tests, 20 μm, this results in approximately a ±5% difference in Young's moduli and approximately a ±2% difference in yield strengths. However, as the effective length of a beam increases compared to its width, the percentage error decreases sharply and the spread of data is expected to decrease as shown in Fig. 9.

In addition to the measurement of the lengths of beams, the measurement of widths and thicknesses can also produce error. In particular, the cross sections of the beams do not remain exactly rectangular because of lateral etching by the wet etchant. Consequently, the error involved in the measurement of widths of beams by SEM could be as large as 1.0 μm. This also leads to another ±5% in uncertainty in Young's moduli and yield strengths for a typical beam width of about 20 μm. The variation in the thickness of thermal SiO₂ was less than ±1%, whereas the thickness of LTO and Au samples can vary by ±1% due to the nature of the deposition techniques employed. This could lead to an additional ±3% difference in Young's moduli and ±2% in yield strengths for LTO and Au thin films. A linear combination of all possible measurement errors produces a total uncertainty in Young's moduli of ±10% for wet thermally grown SiO₂ and ±13% for LTO and Au. For the yield strengths of the Au films the total uncertainty is ±9%. Returning to Fig. 9, the scatter of the measured values of Young's modulus for the thermal oxide falls within the predicted ±10% range.
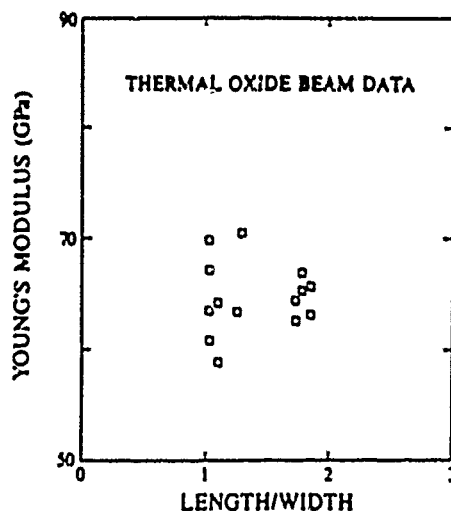
B. Measurement of yield strength and strain hardening

Besides elastic properties, Table I also lists values of hardness and yield strength for the two oxides and Au. Using large beam deflections, the strength of the thermal oxide was measured to be greater than 1.0 GPa. However, since no yielding was noticed, an upper limit on the yield strength of the oxides could not be deter-

FIG. 9. Measured Young's moduli for thermal SiO₂ as a function of the ratio of beam length to width.

mined using the cantilever beam technique. The hardness of the thermal oxide and the LTO were measured to be 12 and 10 GPa, respectively.

For the Au films both yield strength and hardness were determined as shown in Table I. The hardness of 1.2 GPa is approximately 4.6 times the yield strength of 0.26 GPa. This ratio of hardness to yield strength is greater than the more typical value of 3.2, which was predicted by Bhattacharya and Nix in recent finite element studies of indentations.[17] The larger ratio of 4.6 suggests that slightly different plastic behavior is being measured in the two cases. One possible explanation is that the microbeam technique measures a yield strength at a lower plastic strain than 0.2%, which is typically used in comparing hardness to yield strength. Measuring a material's strength at a lower plastic strain would produce a lower yield strength and a higher ratio of hardness to yield strength as observed. Compared to bulk plastic properties, though, the yield strength of the Au thin films is noticeably higher.[18] The refined structure and grain size of the Au thin film, compared to a bulk sample of Au, is thought to be the cause of this difference.

To study the behavior of material following yielding in the microbeam deflections, Fig. 10 presents a large deflection of a Au beam along with theoretical predictions from Fig. 4. As described earlier, an initial linear section indicates purely elastic deformation while deviations from linearity indicates the onset of plastic deformation of the beam. One should notice that while the experimental curve falls between the two theoretical curves (for an elastic–plastic material and a parabolic strain hardening material) it is closer to the elastic–plastic curve. Based on tensile tests of bulk samples of

FIG. 10. Net load and net displacement measurements for a large deflection of a Au cantilever microbeam.

gold,[19] the experimental curve was expected to show higher rates of strain hardening and actually reside above the parabolic strain hardening curve. Thus the relative position of the experimental curve in Fig. 10 suggests that little strain hardening occurs in the Au beams. One possible reason for this behavior is the large surface area to material volume ratio for the microbeams. Compared to bulk samples, the beams have more surface area per unit volume. Consequently, dislocations have a greater chance of gliding to a free surface and leaving the sample. This, in turn, lowers the degree of dislocation interaction for a given amount of deformation and lowers the rate of strain hardening. However, a word of caution is needed. The theoretical predictions in Fig. 10 depend on the measured modulus and yield strength of the Au microbeams, and changing the yield strength used to predict the theoretical curves can shift their position relative to the experimental curve in Fig. 10. Consequently, conclusions regarding the strain hardening behavior of the Au thin films that are based on Fig. 10 must be made with caution.

For most of the unloading of the Au beams, the curves are linear and the reverse deflection is thought to be elastic. However, the initial section of the unloading curves in Figs. 7, 8, and 10 is nonlinear, and all other tests of Au beams show similar nonlinearities. This deviation from linearity is thought to be caused by the indenter tip sliding on the microbeam. At the start of unloading, the indenter is believed to slide out along the beam, thereby permitting the beam deflection to remain relatively constant while the load decreases. Once sliding stops, the unloading curve becomes linear as expected.

In Fig. 10, which shows a large deflection of a Au beam, the nonlinearity in the unloading section of the curve is most significant. For this test the loading position on top of the beam can be seen with an SEM because large loads were applied. Furthermore, a troughlike shape can be detected leading away from the initial loading position, out along the beam. Although the trough is difficult to resolve clearly, it is thought to mark the slip of the indenter tip on the microbeam. Such slip, which is enhanced by the rounded, nonideal tip of the triangular indenter, is highly plausible at the start of the unloading because the curvature of the beam is at its maximum. The slight decrease in the unloading slope compared to the loading slope in Fig. 10 also supports this hypothesis. Assuming the elastic modulus remains constant through the whole test, the effective length c for unloading must have increased by 2.1 μm to cause the decrease in the load-deflection slope from loading to unloading. This displacement matches well with the 2.2 μm length of trough seen on top of the beam.

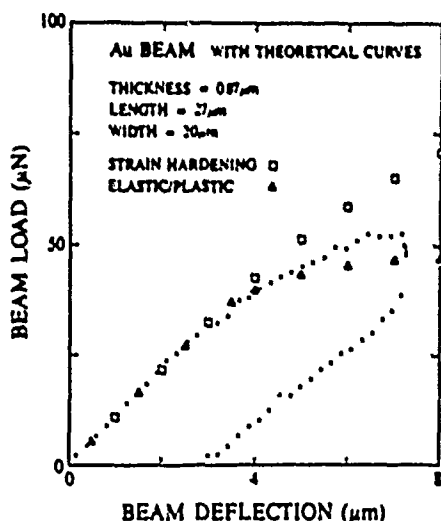In contradiction to the above argument one could reason that the indenter's tip slipped during the loading

of the beam and not during the unloading. One may even suggest that this slip is at least partially responsible for the reduction of the slope of the loading curve. However, if this were the case, then the permanent deflection of the beam (which is marked by the offset between the loading and unloading deflections at zero load) would not equal the deviation from the linear loading as shown in Fig. 10. Thus, even though / .e evidence is not conclusive, it does suggest that the tip of the indenter slips on the Au beams at the start of the unloading and causes a nonlinearity in the unloading curve.

C. Additional deflections of the microbeams and the tip of the Nanoindenter

Although the simple beam theory and the shape of the load-deflection curves are in agreement, there are several other deflections associated with the microbeams and the tip of the Nanoindenter that need to be discussed. The first of these is caused by the fact that the beam is wide compared to its thickness and that the beam is point loaded. As the diamond tip contacts a beam at its center and applies a load, the beam tends to bend across its width. The center of the beam bends downwards relative to the edges of the beam. Since this curvature would not have resulted from a uniform loading across the beam's width, and since Eq. (1) is actually derived for such a loading, the added deflection due to this curvature needs to be considered. This is done in Appendix B. It amounts to approximately 2% of the deflection given by Eq. (1) for a beam in which the ratio of the width and the effective length is about unity. As the ratio of effective length to width increases, the relative amount of additional deflection decreases markedly. However, the additional displacement has been subtracted in the elastic analysis to more accurately determine the elastic moduli of the microbeams.

The second additional deflection is that of the indenter tip pushing into the top of the microbeams, both elastically and plastically. To determine the elastic portion of the indentation one can consider the compliance of the material under the indenter tip. For Au and for the oxides tested, this compliance is less than $50 \mu m/N$. For the Au and oxide beams, the measured compliances are approximately 1000 times greater. Therefore, the elastic indentation of the beams is less than 0.1% of the gross elastic deflections of the beams and can be ignored.

The plastic component of an indentation of the microbeams, however, may not be simply ignored. For soft materials such as Au, the plastic portion of an indentation is significantly greater than the elastic portion. In Fig. 11 one can see that although the plastic indentation depths (0 to 40 nm) are small compared to both the elastic and plastic deflections of the Au beams, they are measurable for the loads applied to the microbeams (0–80 μN). Thus the plastic indentation depths were sub-
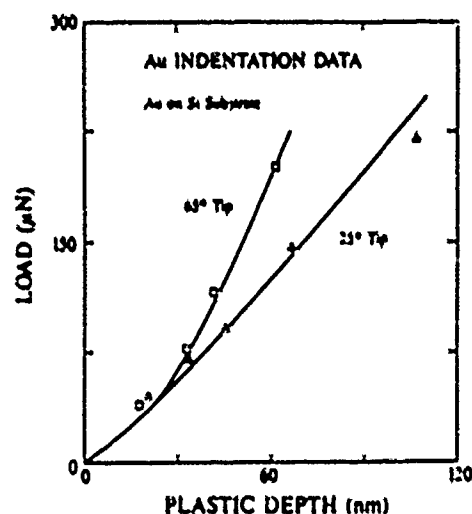

FIG. 11 Load and plastic depth measurements for submicron indentations of a 0.87 μm Au film on a Si substrate. Two different diamond tips were used.

tracted from the total measured deflections of the beams. For very hard materials such as SiO_2, the plastic component of the indentation is similar in magnitude to the elastic portion and therefore can be easily ignored. The lack of an offset between the loading and unloading curves in Figs. 5 and 6 verifies this statement.

VII. CONCLUSION

Young's moduli and yield strengths of Au and SiO_2 thin films have been measured using a Nanoindenter in conjunction with cantilever microbeam samples. The measurements were made on single-layer, micron-thick beams that were fabricated using conventional integrated circuit (IC) fabrication techniques. The free standing beams were mechanically deflected by a Nanoindenter. From the load-deflection data acquired during bending, the local elastic and plastic properties of the thin films were determined without ever directly handling the samples. The moduli measured for Au, LTO, and thermal oxide and the yield strength measured for Au are in agreement with previously published results for these materials.

ACKNOWLEDGMENTS

poration, and the Gifts Funds from the Center for Integrated Systems at Stanford University.

APPENDIX A: AN ESTIMATION OF THE RESIDUAL STRESSES IN Au BEAMS AFTER FABRICATION

The curvature of Au beams in an as-fabricated, freestanding state (Fig. 2) strongly suggests that residual stresses are present when the films are deposited onto Si substrates. These as-deposited stresses can develop for a variety of reasons, but the main concern is whether or not stresses remain in the thin films when the substrates are etched away from below the beams. To answer this question definitively, one needs to know the source of the as-deposited stresses, prior to etching away the substrate. However, investigating the possible sources of stresses is beyond the scope of this present study. Thus we can only conjecture plausible sources and then determine how the removal of the substrates affects the as-deposited stresses.

Although there are a variety of sources for residual stresses in as-deposited thin films, they can be grouped roughly into three types: lattice mismatch, thermal mismatch, and growth mismatch. The first two sources are not applicable to the Au samples prepared here because they were neither grown epitaxially nor deposited at elevated temperatures. Only the last type of source, growth mismatch, can account for the bending of the Au microbeams.

When a metal is grown on a substrate in polycrystalline form, stresses can arise in the film through grain growth or vacancy annihilation. Both processes act to densify the film. Since the substrate constrains the movement of the film, a state of tension can develop in the film. If the tensile stresses are uniform across the thickness of the film, no bending results when the substrate is removed. However, if the stresses vary across the thickness of the film, then the free-standing beams will bend. In the case of the Au beams, the bottom layers of atoms may have had more time and energy to rearrange than the top layers of atoms. Thus more tension may have developed in the bottom layers compared to the top layers. When the substrate is removed, this stress state can cause a downward bending of the beams. Yet, for a linear variation of growth stresses over the film's thickness, the uniform contraction, and the bending of the beams, will remove the residual stresses present in the as-deposited films. Only a nonlinear or discontinuous variation in the as-deposited stresses can produce stresses in the free-standing beams after they bend.

The as-deposited stress state that would produce the largest residual stress in the free-standing beams is significant tension in the Cr underlayer and little or no stress in the Au film. Such a stress state could develop if the thin Cr under layer densifies through the growth of

grains or the annihilation of vacancies while the Au film remains as-deposited. The possible stresses in the Cr under layer and the Au film resulting from such an as-deposited state can be approximated by measuring the radii of curvature of the free-standing beams.[13] Using such a technique for the Au beams shown in Fig. 2, the stress in the Cr under layer was calculated to be approximately 1.3 GPa. But, for the gold film, the stresses are much lower. As a rough approximation, they scale according to the relative thickness of the Cr and Au 1 to 200. This suggests that the maximum possible stress in the Au beams is less than 7 MPa. Considering that the yield stress of the Au is almost 50 times greater, and that the standard deviation of the measured yield strength is almost 4 times greater, the possibility of residual stresses in the free-standing beams affecting measured strengths can be ignored.

APPENDIX B: AN ESTIMATION OF THE CURVATURE EFFECT ABOUT A POINT LOADING ON A CANTILEVER BEAM

When a force is applied at a point on a cantilever beam with a finite width and length, both of which are much greater than the thickness of the beam, curvature is generated across its width. This curvature is in addition to the curvature along the beam's length. Using the method of superposition and plate theory,[24] the deflection of a cantilever beam due to the curvature across its width was calculated separately and compared to the deflection given by Eq. (1).

A schematic diagram of the test configuration used here is shown in Fig. 12, and this is approximately divided into two individual loading configurations: one is just the beam bending condition (see Fig. 13) for which the deflection solution is simply given by Eq. (1). The other configuration is a point force opposed by a uniformly distributed line load in the opposite direction on an infinite strip with simply supported edges (see Fig. 14). The addition of the deflection solutions for these two configuration is assumed to be equivalent to the total deflection solution for the original test configuration, Fig. 12.

The deflection equation for the infinite strip case is derived as follows. When there is no load acting on the
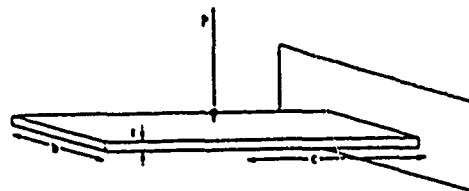

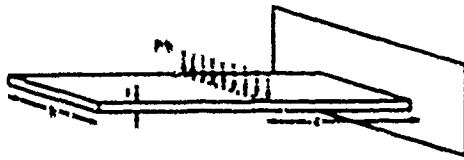FIG. 12. Schematic of the actual cantilever microbeam test geometry

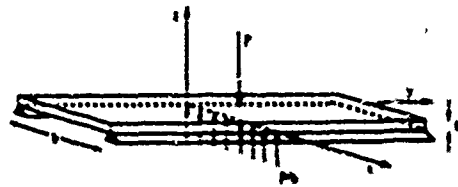FIG. 13. Schematic of linear loading of a thin cantilever microbeam for which Eq. (1) is derived.

FIG. 14. Schematic of a combined loading of an infinitely long plate linear loading on bottom of plate, point loading on top of plate, and simple side supports.

plate, the differential equation governing its deflection, w, is given by

$$\frac{\partial^4 w}{\partial x^4} + 2\frac{\partial^4 w}{\partial x^2 \partial y^2} + \frac{\partial^4 w}{\partial y^4} = 0.$$

The solution of this equation is assumed to be in the following form:

$$w = \sum_{m=1}^{\infty} Y_m \sin\frac{m\pi x}{b}.$$

The boundary condition at the midpoint through the thickness of the beam ($y = 0$) is

$$\left(\frac{\partial w}{\partial y}\right)_{y=0} = 0.$$

Then, the deflection w is

$$w = \sum_{m=1}^{\infty} B_m\left(1 + \frac{m\pi y}{b}\right)e^{-\frac{m\pi y}{b}}\sin\frac{m\pi x}{b}.$$

Now, the constant B_m can be determined if the load distribution along x axis is given. A Fourier expansion of load at $y = 0$ is

$$V_y = V_0(x) = \sum_{m=1}^{\infty} V_m \sin\frac{m\pi x}{b},$$

where

$$V_m = \frac{2}{b}\int_0^b V_0(x)\sin\frac{m\pi x}{b}dx$$

$$= -\frac{2P}{b}\left(\sin\frac{m\pi}{2} + \frac{1}{m\pi}(\cos m\pi - 1)\right).$$

Then,

$$V_0(x) = -\frac{2P}{b}\sum_{m=1}^{\infty}\left(\sin\frac{m\pi}{2}\right.$$
$$\left. + \frac{1}{m\pi}(\cos m\pi - 1)\right)\sin\frac{m\pi x}{B}.$$

Since the load is equally divided between two halves of the plate, the shearing force with respect to the y axis should be

$$Q_y\big|_{y=0} = -D\frac{\partial}{\partial y}\left(\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2}\right)_{y=0} = \frac{V_0(x)}{2},$$

where

$$D = Et^3/12(1 - v^2).$$

The constant B_m that satisfies the above boundary condition is

$$B_m = \frac{(P/b)[\sin(m\pi/2) + (1/m\pi)(\cos m\pi - 1)]}{2D(m\pi/b)^3}$$

Thus,

$w(x,y)$

$$= \sum_{m=1}^{\infty}\frac{P[\sin(m\pi/2) + (1/m\pi)(\cos m\pi - 1)]}{2bD(m\pi/b)^3}$$

$$\times\left(1 + \frac{m\pi y}{b}\right)e^{-\frac{m\pi y}{b}}\sin\frac{m\pi x}{b}.$$

At $y = 0$,

$w(x,0)$

$$= \sum_{m=1}^{\infty}\frac{P[\sin(m\pi/2) + (1/m\pi)(\cos m\pi - 1)]}{2bD(m\pi/b)^3}$$

$$\times\sin\frac{m\pi x}{b}.$$

Since w converges rapidly for m greater than 1, the deflection in the middle of the plate at $y = 0$ is approximately

$$w\left(\frac{b}{2}, 0\right) \approx \frac{Pb^3(1 - 2/\pi)}{2D\pi^3}$$

Therefore, the ratio of curvature deflection to simple beam deflection is given by

$$\frac{w(b/2, 0)}{w_{beam}} \approx \frac{3(1 - 2/\pi)(1 - v^2)}{2\pi^3}\left(\frac{b}{c}\right)^3$$

$$\approx 1.7 \times 10^{-2}\left(\frac{b}{c}\right)^3.$$

In summary, the point load on the cantilever beam produces an extra deflection not predicted by the simple beam bending equation shown previously. This relative additional deflection is proportional to the cube of the ratio of the width and the length of the beam. For the lengths and widths tested, the extra deflection is less than 2.0% of the total beam deflection measured by the Nanoindenter.

REFERENCES

[1] M. F. Doerner and W. D. Nix, J. Mater. Res. 1, 601 (1986)

[2] J. Pethica, R. Hutchings, and W. C. Oliver, Philos. Mag. A 48, 593 (1983).

[3] J. D. Finegan and R. W. Hoffman, Transactions of the 8th National Vacuum Symposium (Pergamon, New York, 1961), p. 935.

[4] A. K. Sinha and T. T. Sheng, Thin Solid Films 48, 117 (1978)

[5] E. J. McInerney and P. A. Flinn, in Proceedings of the 20th Annual International Reliability Symposium, IEEE Electron Devices and Reliability Societies (1982), p. 264.

[6] J. W. Beams, J. B. Freazeale, and W. L. Bart, Phys. Rev. 100, 1657 (1955)

[7] T. Tsakalakos and J. E. Hilliard, J. Appl. Phys. 54, 734 (1983)

[8] R. W. Hoffman, C. O. Andeen, and C. W. Hagerling, Proc. 7th Int. Vacuum Congr., Vienna 2, 1769 (1977)

[9] K. E. Peterson and C. R. Guarnieri, J. Appl. Phys. 50, 6761 (1979)

[10] S. P. Timoshenko and J. M. Gere, Mechanics of Materials (Van Nostrand, New York, 1972), p. 179.

[11] R. J. Jaccodine and W. A. Schlegel, J. Appl. Phys. 37, 2429 (1966).

[12] C. A. Neugebauer, J. Appl. Phys. 31, 1096 (1960)

[13] A. Kinibera, S. Baba, and N. Matuda, Thin Solid Films 141, 229 (1966)

[14] CRC of Chemistry and Physics, edited by R. C. Weast (CRC, Boca Raton, FL, 1985), p. F-54.

[15] W. Koster and H. Franz, Metall. Rev. 6(21), 1 (1961)

[16] H. B. Huntington, The Elastic Constants of Crystals (Academic, New York, 1983), p. 112.

[17] A. Bhattacharya and W. D. Nix, Int. J. Solids Struct. (submitted for publication).

[18] Metals Handbook, Properties and Selection of Metals (ASM, Metals Park, OH, 1980), 9th ed., Vol. 2, p. 714.

[19] Y. Nakada, U. F. Kocks, and B. Chalmers, Trans. AIME 230, 607 (1964)

[20] S. P. Timoshenko and K. Woinowsky-Krieger, Theory of Plates and Shells (McGraw-Hill, New York, 1940), p. 167

942

J. Mater. Res., Vol. 3, No. 5, Sep/Oct 1988

INJECTION OF DISLOCATIONS INTO STRAINED MULTILAYER STRUCTURES

by

J.P. Hirth
Department of Mechanical and Materials Engineering
Washington State University
Pullman, WA 99164-2920

ABSTRACT

The problem of defect injection into strained multilayer structures (superlattices) is addressed. Equilibrium critical thicknesses are determined in a Volterra dislocation model, extending earlier work to the large misfit regime. The corresponding problem of spreading of a threading dislocation is treated. Finally, the nucleation and growth processes for creating misfit dislocations are considered.

INTRODUCTION

The work of Guinier [1] and Preston [2] indicating elastic strains associated with GP precipitates established the concept of coherency strains connected with coherent interfaces, including those for isolated precipitates and those for planar interfaces of oriented overgrowths on layer structures. The theoretical work of Frank and Van der Merwe [3] showed that the coherency strains could be relieved by the presence of misfit dislocations. Misfit dislocations were soon thereafter verified by etch pitting [1] or by transmission electron microscopy [5,6]. Since then, there have been a large number of observations of misfit dislocations as reviewed by Matthews [7] for early work.

With the demonstration of the potential of strained multilayer structures for electronic applications such as quantum well structures [8,9] there has been renewed interest in the interface structure. The present work is restricted to oriented single or multilayers with planar interfaces: pertinent to the proposed electronic applications.

An oriented overgrowth layer is said to have an epitaxial or endotaxial orientation with respect to the substrate layer and the interface is identified by the crystallographic habit places of the two adjoining crystals. If there is one-to-one atom matching at the interface the interface is coherent or, in the recent alternate terminology in condensed matter physics, commensurate. If the strain is relieved by the introduction of misfit dislocations, the interface becomes semicoherent. The initial terminology [8] for an alternating sequence of strained coherent layers was a strained superlattice, but this choice was unfortunate because of confusion with the with the crystallographic superlattices that occur for ordered compounds. The alternative description as a strained layer structure is followed

here.

Thin strained layers tend to be stable in the coherent state, both thermodynamically and kinematically, while thicker layers become unstable with respect to misfit dislocation injection. The consideration of the critical instability condition is the major thrust of the present work. There are four aspects of the instability problem: (i), the equilibrium interface state, (ii), partial equilibrium states with incomplete dislocation arrays, (iii), the spreading of dislocations that thread through the layers, and (iv), the injection and spreading of dislocations by nucleation, multiplication and motion. We briefly discuss the first three aspects of the problem and concentrate on the fourth, which is the one of most practical interest in limiting the performance of devices based on strained multilayers.

DISLOCATION BEHAVIOR

Figure 1 illustrates a coherent interface, an interface containing a misfit location, an interface containing a misorientation dislocation, and a combined dislocation with both misfit/misorientation character, so designated to distinguish it from a mixed screw/edge dislocation. As found by Giradmann [10] and others [7], many of the dislocations formed to relieve misfit have combined character and so contribute to misorientation as well. For simple cases such as two fcc crystals meeting at a plane of high symmetry, the nature of the dislocations, perfect or partial dislocations belonging to one or the other bulk lattice, are easy to comprehend. In less simple cases, two lattices are related by a CSL correspondence lattice [11] (a lattice of crystal positions common to both bulk lattices). The reciprocal of this lattice is the DSC lattice and the Burgers vectors of interface dislocations are perfect or partial ID's (interface dislocations) belonging to the DSC lattice: the perfect

and partial Burgers vectors of the two bulk lattices also belong to the DSC lattice.

If the dislocation is not a perfect bulk dislocation, it will have an interface fault associated with it, as illustrated for the simple case of a twin boundary in Figure 2. Ledges at interfaces, because of differences in lattice spacing in the two lattices, perforce have an associated dislocation called a step translational dislocation, with ID character. Fault translational dislocations, which are also ID's, separate interface regions of different fault nature.

There are a few other types of dislocations at interfaces. Continuous misorientations across an interface, e.g., at a first order twin boundary, can be described in terms of intrinsic ID's [12, 13]. The coherent strain field can be considered to arise from a continuous distribution of infinitesimal dislocations. Finally, small added elastic strains, associated with dilatations normal to faults for example, can lead to small dislocations with translational character. In neither of the two latter two cases do the Burgers vectors belong to the DSC lattice. Detailed discussions of these dislocations are presented elsewhere [12-15]. Here, our primary emphasis will be on misfit or combined dislocations with perfect bulk Burgers vectors.

The original analysis of Frank and van der Merwe and continued analysis by the latter author [15, 16] are in terms of a description of the dislocations by the Frenkel-Kontorova model. Later analyses have been made in terms of the Volterra model [17]. Other Peierls-type models are also possible, with either the standard sinusoidal form [18, 19] or with a parabolic form [20]. We emphasize that, for single dislocations, the long-range field in all cases converges to the Volterra linear elastic result. All of these models give artificial descriptions of the near-core regions: none, for example, contain the core

dilatation that is a nonlinear elastic feature of atomic simulations of dislocations [21]. Also, there is no first-principles physical calculation of core structure, although atomic simulations have been made using an empirical potential for silicon [22]. Hence, we follow here the simple Volterra model for the dislocations. Since most calculations involve dislocation interactions, this model should be accurate when the dislocation spacing is large, say ten or more atomic spacings, compared to core dimensions. For more closely spaced dislocations, core interactions and core relaxations could make the Volterra model physically unrealistic, in which case one of the other models could be superior.

Anisotropic elastic descriptions are available [23-25] and have been used in some calculations. However, in most cases, the complexity of such calculations, involving numerical methods, makes the use of anisotropic elasticity cumbersome and we shall restrict the present analysis to isotropic elasticity. A calculation [26] of the thermodynamic critical thickness for the Ga-As-P multilayer system by both methods showed that the anisotropic result differed from the isotropic one by about 25 percent, so errors of this order can be anticipated in the use of isotropic elasticity.

Image forces associated with elastic inhomogeneity [27] can also influence results, driving misfit dislocations off the interface and into the softer bulk material, for example. This effect becomes important when the elastic constants of the bulk crystals differ by more than about 50 percent. In most cases of interest for electronic applications, the elastic constants do not differ by this amount and the inhomogeneity effect can be neglected with less resultant error than the neglect of anisotropic elastic effects. A counter example is the Nb/Al_2O_3 case where the image effects drive misfit dislocations five atom spacings into the softer Nb crystal [27, 28].

Finally, diffusional effects influence the forces on dislocations. The graded concentrations produced near an interface by interdiffusion distributes and lowers the coherency strain near the interface and hence changes the elastic force on misfit dislocations [7, 29, 30]. Flux divergences at interfaces growing by diffusion and/or vacancy annihilation fluxes lead to osmotic forces (climb forces) on dislocations that can be large compared to elastic forces [31]. These effects have not been considered quantitatively in any treatment of the injection or motion of misfit dislocations and represent an opportunity for further work.

INTERFACE EQUILIBRIUM

1. Elastic Stresses

As a simple example of the thermodynamic limit for instability we consider the example in Figure 3 of two cubic crystals meeting on (1 0 0) habit places.

In a multilayer structure the stresses are uniform except near free surfaces. For a layer in the multilayer structure of Fig. 3, coherency gives rise to equal biaxial stresses $\sigma_{11} = \sigma_{22} = \sigma$ and strains $\epsilon_{11} = \epsilon_{22} = \epsilon$. The boundary conditions $\sigma_{33} = 0$ then leads to the result

$$\sigma = \epsilon\epsilon \tag{1}$$

where

$$\epsilon = 2\mu(1 + \nu)/(1 - \nu) = 2\mu\kappa \tag{2}$$

with μ the shear modulus, ν Poisson's ratio, and κ an abbreviated Poisson factor. Since no net force can act on the system, the stresses partition in the two layers so that

$$\sigma_A h_A + \sigma_B h_B = 0 \tag{3}$$

with h the layer thickness. The misfit strains must also add to remove the total misfit.

$$\epsilon_0 \equiv \Delta a / <a> = \epsilon_A + \epsilon_B \tag{4}$$

where a is the lattice parameter, $\Delta a = a_B - a_A$ and $<a> \cong a$ since in all cases of interest $\epsilon_0 < 0.01$. The above equations combine to give the result

$$\epsilon_1 = \epsilon_0 h_B c_B / (h_A c_A + h_B c_B)$$

$$\epsilon_B = \epsilon_0 h_A c_A / (h_A c_A + h_B c_B) \tag{5}$$

At the free edge surfaces, the normal stresses must be released to satisfy the free surface boundary condition. This can be accomplished by imposing edge tractions to cancel the in plane stresses, Fig. 1. The resulting stress field can be calculated by superposing results for line forces acting on half-spaces [32]. The result for the interface shear stress at the central interface among N layers is a sum which to first order gives

$$\sigma_{12} = 2\epsilon_0 h^2 / \pi (h^2 + x^2) \tag{6}$$

This stress falls off with distance x into the interface, becomes negligibly small when $x > 4h$, and has a maximum value at the surface of

$$\sigma_{12} = 2\epsilon_0 / \pi \tag{7}$$

The peak interface shear is thus similar in magnitude to the uniform normal stress, Eq. (4).

For thin bilayers there would also be stresses associated with the relaxation of bending moments [33], but these are negligible for the multilayer case. For multilayers coherently affixed to a substrate, the elastic field is also modified. For example, if a multilayer of A and B were deposited on a thick B substrate the strain fields would be $\epsilon_A = \epsilon_0, \epsilon_B = 0$ instead of Eq. (5).

Misfit Dislocation

For one simple fcc {100} case, the long-range strain field can be completely removed by edge misfit dislocations $\frac{1}{2}$ < 110 > lying in the interface, Fig. 5, an array sometimes observed [7]. Counting of lattice planes for Fig. 5 shows that the misfit dislocation spacing is

$$\lambda \cong h < n > /\Delta n = b/\epsilon_n \qquad (8)$$

with ϵ_n and $< n >$ defined as in Eq. 1. If λ/h is close to an integer value, the small difference can be accommodated by a set of translational D's with no net b when summed over the habit plane. If λ/h is not close to an integer value, then a sequence of alternating spacings λ_i is required such that $< \lambda_i > = b/\epsilon_n$. The latter complication becomes more important for small λ.

In determining the equilibrium arrays, one could work with global energies and convert the previous coherency strain fields to strain energies. As with many dislocation problems, such a procedure is cumbersome and a simpler method is to determine the Peach-Koehler thermodynamic force on the dislocation and integrate it to determine _local_ interaction energies. In order to illustrate the procedures, which would involve lengthy numerical sums for the multilayer case, we consider the case of a single layer of thickness h_A between two B layers with thicknesses $h_B >> h_A$ and with $\epsilon_A = \epsilon_B = \epsilon$. The misfit dislocation array can then be considered to arise by the reversible process illustrated in Figure 6.

As shown in Figure 6, a set of dislocation dipoles of infinitesimal separation and with dipole spacing λ is imagined to form. The dislocations of set II are then separated first on the plane $y = 0$ to a position $x = r_e$ and then separated at $x = r_e$ to a position $y = h_A$.

Since the spacings of set II dislocations do not change in this process there is no self work done by dislocations in set II in the overall process (equivalently the Peach-Koehler force is zero among set II dislocations). However the motion of the set II dislocations results in reversible work corresponding to the forces arising from set I dislocations and from the coherency stresses. The relevant components of the Peach-Koehler force are

$$\frac{F_x}{l} = -\sigma_{xy} b_x \tag{9}$$

and

$$\frac{F_y}{l} = \sigma_{xx} b \tag{10}$$

where $b_x = b = \frac{1}{2} < 110 >$.

In the evaluation of σ_{xy} and σ_{xx}, it is convenient to work with reduced distances $X = x/\lambda$ and $Y = y/\lambda$. The stresses of the set I dislocations in the configuration of Figure 6 are well known [31] and have the components

$$\sigma_{xy} = \sigma_0 \sin 2\pi X (\cosh 2\pi Y - \cos 2\pi X - 2\pi Y \sinh 2\pi Y) \tag{11}$$

$$\sigma_{xx} = -\sigma_0 [2 \sinh 2\pi Y (\cosh 2\pi Y - \cos 2\pi X - 2\pi Y (\cosh 2\pi Y \cos 2\pi X - 1)] \tag{12}$$

where

$$\sigma_0 = \mu b / 2\lambda (1-\nu)(\cosh 2\pi Y - \cos 2\pi X)^2 \tag{13}$$

The work of interaction is the same for each dislocation in set II. Hence the forces of Eqs. (9) and (10) need only be integrated for one dislocation. The result is the interaction and coherency work per dislocation dipole pair and can be divided by two to get the work per misfit dislocation. Moreover, one should actually consider the perpendicular set of misfit dislocations of Fig. 5 to form at the same time. Analogous to the result for the biaxial

coherency stress, added work must be done for set II dislocation to compensate for the displacement u_z produced by the cross-grid set of dislocations parallel to z and this will increase the work by a factor $(1 + \nu)$. Thus, the energy per dislocation using our work done by all thermodynamic forces is

$$\frac{W}{L} = \frac{(1 + \nu)}{2} \int_0^{h_A} \frac{F_y}{L} dy + \frac{(1 + \nu)}{2} \int_0^{x_c} \frac{F_z}{L} dz$$

$$= \frac{\lambda(1 + \nu)}{2} \int_0^{h_A/\lambda} (\sigma_{rr}^D + \sigma_{rr}^C) b d Y - \frac{\lambda(1 + \nu)}{2} \int_0^{X_c} \sigma_{xy}^D b d X \qquad (11)$$

where the superscri:, s D and C indicate, respectively, the the dislocation fields of Eqs. (11) and (12) and the coherency fields of Eqs. (1) and (5). When h_A is less than a critical value h_c, W/L is positive and the coherently strained state is stable. When h_A is greater than h_c, W/L is negative and the misfit dislocation state is stable. We consider solutions to Eq. (11) first in the two limiting cases $h_c \ll \lambda$ and $x_c = 1/2$ and then for the general case.

3. Equilibrium Arrays and Critical Thicknesses

The configurations in the two limiting cases are shown in Fig. 7. In the limit $h_c \ll \lambda$, the stress fields of Eqs. (11) to (13) reduce to those for a single edge dislocation and the equilibrium configuration is the lone dislocation dipole, 45° result [34]. This is the case treated by Matthews and Blakeslee [7, 17] and we term it the MB result. The total energy from Eq. (14) is

$$\frac{W}{L} = \frac{\mu b^2 \kappa}{4\pi} \left[\ln \left(\frac{\sqrt{2} h_A}{r_o} \right) - \frac{4\pi h_A}{\lambda} \right] \qquad (15)$$

where $r_o \cong b/3$ is the core cutoff radius typical for covalent crystals [18], the outer cutoff radius is the dipole separation $\sqrt{2}h$, the first term in square brackets is the dislocation interaction energy and the second term is the work done by the coherency stress in forming

the dipole. Normalized to unit area of interface, the total energy is the above result multiplied by the total lengths of dislocation in the square grid, $2/\lambda$, or

$$\frac{W_T}{l} = \frac{\mu b^2 \kappa}{2\pi\lambda} \left[\ln\left(\frac{3\sqrt{2}h_A}{b}\right) - \frac{4\pi h_A}{\lambda} \right] \qquad (16)$$

Thus, the MB result for the critical thickness where $W_T/l = 0$ is

$$h_c / \ln(3\sqrt{2}h_c/b) = \lambda/4\pi = b/4\pi\epsilon_0 \qquad (17)$$

For a multilayer with $h_1 = h_B = h$ in the MB limit, the dislocation interaction energy would be the same but the coherency stress in the A layer would be one-half as large. The result would thus be of the form of Eqs. (15) and (16) with the last term in square brackets one-half as large. The critical thickness would be

$$h_c / \ln(3\sqrt{2}h_c/b) = \lambda/2\pi = b/2\pi\epsilon_0 \qquad (18)$$

For the single A layer, as h increases, the glide equilibrium position X_e approaches $1/2$ and equals $1/2$ for $h > 0.215\lambda$. In this limit, the test misfit dislocation configuration is that of Fig. 7b. In this limit, a convenient integration path for Eq. (11) is to first integrate over X to $X_e = 1/2$ and then integrate over Y to h_A/λ at $X = X_e$. For the first integral, σ_{xy}, Eq. (11), reduces to

$$\sigma_{xy} = \frac{\mu b}{2\lambda(1-\nu)} \frac{\sin 2\pi X}{(1 - \cos 2\pi X)} \qquad (19)$$

For the second integral, σ_{xx}, Eq. (12), reduces to

$$\sigma_{xx} = -\frac{\mu b}{2\lambda(1-\nu)} \frac{(2\sinh 2\pi Y + 2\pi Y)}{(\cosh 2\pi Y + 1)} \qquad (20)$$

The resulting total energy is

$$\frac{W_T}{l} = \frac{\mu b^2 \kappa}{4\pi\lambda} \left\{ \ln[\cosh\alpha + 1] - \ln[\cosh\gamma - 1] + \alpha\tan(\alpha/2) - 1 - 4\alpha \right\} \qquad (21)$$

where $\alpha = 2\pi h_A/\lambda$ and $\gamma = 2\pi b/\lambda$. Thus, for this case of $X_e = 1/2$, $Y_e = h_A/\lambda$ and for a given λ, the value of h_{e_1} or equivalently α_{e_1} is determined for the case $W_T/L = 0$, i.e.,

$$\ln[\cosh\alpha_c + 1] + \alpha_c \tanh(\alpha_c/2) - 4\alpha_c = \ln[\cosh\gamma - 1] + 1 \tag{22}$$

For the general case where $0 < X_c < 1/2$, the resulting total energy is

$$\frac{W_T}{L} = \frac{\mu b^2 \kappa}{4\pi\lambda}\left(\ln[\alpha\sinh\alpha] - \ln[\cosh\gamma - 1] - \ln\right) \tag{23}$$

Here, the glide equilibrium position X_e or equivality $\eta_e = 2\pi X_e$, for a given α_{e_1} is determined from

$$\cos\eta_e = \cosh\alpha_e - \alpha_e\sinh\alpha_e \tag{24}$$

The critical value α_c is determined for the case $W_T/L = 0$ from the relation

$$\ln[\alpha_c\sinh\alpha_c] - 4\alpha_c = \ln[\cosh\gamma - 1] \tag{25}$$

For the multilayer case, the fields of sets of dislocations such as those in Fig. 7 must be summed. The details are presented elsewhere [35] and are beyond the scope of this overview. However, we present the result for the sum of terms of the type of Eq. (21) to show the form of the result. In the multilayer case, this limiting form $X_e = 1/2$ is more favored than for the single layer case and applies for almost all ranges of λ.

We define summed functions

$$p(m) = \ln\sinh(m\alpha/2) + (m\alpha/2)\coth(m\alpha/2)$$

$$q(m) = \ln\coth(m\alpha/2) - m\alpha/\sinh(m\alpha)$$

where m is an integer. The total energy per layer for an n multilayer is then

$$\frac{W_T}{L} = \frac{\mu b^2 \kappa}{2\pi\lambda}\left\{\frac{1}{(n+1)}\sum_{j=1}^{n}(-1)^{j+1}jp(n-j+1) + \frac{1}{(n+1)}\sum_{j=1}^{(n+1)/2}(2j-1)q(n-2j+2)\right.$$

$$\left. - \frac{1}{2}[\ln\sinh(\gamma/2) + 1] - \alpha\right\} \tag{26}$$

A comparison of the results from Eqs. (16), (21) and (23) is given in Fig. 8 for a sample case. As can be seen, the multiple dislocation results differ from the M-B result of Eq. (16) when h_c/λ becomes small. This verifies the results of Jesser and van der Merwe [15], although, as discussed in detail elsewhere [35], they treated a constrained glide equilibria case. However, we note that the M-B result does not apply except in the limit $\lambda \gg h$ not because they used the Volterra dislocation approximation as suggested in Ref. [15], but because M-B neglected multiple dislocation interactions. Indeed, as mentioned previously, the parabolic core potential results for the dislocation strain field [15, 20] should agree with the Volterra results except in the core vicinity. The form of Eq. (26) is similar to the parabolic potential results, although the specific sums and integrals differ. The Volterra and parabolic potential results are expected to differ only in the core term (reflected by b or γ in the present results) except in the limit γ or $h \to b$, where the elastic fields would differ.

IV. Partial Equilibrium Arrays

For Eqs. (21), (23) and (26) and their multilayer counterparts, W_T/l, monotonically decreases with decreasing λ to the equilibrium value λ_e of Eq. (8) once $h > h_c$. Once the formation of misfit dislocations becomes thermodynamically favorable, then the formation of the equilibrium cross-grid of misfit dislocations becomes favorable also. For the pure misfit dislocation case, in the absence of frictional forces or other extrinsic effects, there is no constraint to the formation of the equilibrium misfit dislocation array.

However, there are several reasons why complete equilibrium might not be achieved once h exceeds h_c. First, if misfit is relieved by the combined dislocations of Fig. 1d

' stead of the misfit dislocations of Fig. 1b, added dislocation interactions arise that influence the local equilibrium values of h and λ. For the (0 0 1) diamond cubic interface case, the combined dislocations could glide on the system $\frac{1}{2}[0\ 1\ 1](1\ 1\ 1)$, Fig. 9a, to create pure edge, combined dislocations with length $h' = (\sqrt{6}/2)h$. The interaction energy is then of the form in Eq. (15) but with h' replacing $\sqrt{2}h$. Only the misfit component $b_m = \frac{1}{2}[6\ 1\ 0] = (\sqrt{2}/2)b$ would relieve coherency stresses so that for the coherency portion of Eq. (15) b_m would replace b. Hence Eqs. (16) and (17) would be modified to the form

$$\frac{W}{L} = \frac{\mu b^2 \kappa}{2\pi\lambda} \left[\ln\left(\frac{3\sqrt{6}h}{2b} \right) - \frac{2\pi h c_o}{b} \right] \tag{27}$$

and

$$h_c / \ln(3\sqrt{6}h_c/2b) = b/2\pi c_o \tag{28}$$

Hence, the critical thickness would be larger than for the misfit dislocation case.

An analogous qualitative trend holds in the multilayer case, Fig. 9b. For this case there are added interactions for all dislocations and each term in Eq. (26) would be modified. Just as for the single layer case, h_c would be increased for the multilayer case. In the configuration of Fig. 9b, the misorientation dislocation components of the dislocations, between A and B for example, are also repulsive. Thus, unlike the pure misfit dislocation case, for the combined dislocations W/L does not monotonically decrease with decreasing λ to λ_e satisfying Eq. (8) once h exceeds h_c. Hence, there will be a constrained equilibrium value $\lambda_m > \lambda_e$ and a corresponding residual coherency strain present. As h increases, when $h > h_c$, λ_m approaches λ_e asymptotically and the coherency strain vanishes only in this limit.

In a similar manner, when h or $\lambda \rightarrow b$, core-core interactions between dislocations

become important, modifying the interaction energy portion of the various expressions for W/L and hence also for h_c. The parabolic potential results [15], for example, contain this feature. There is little guideline to the form of the core interaction from atomistic calculations because of uncertainties in the interatomic potentials. However, available results [21] suggest that the very short range (one or two atom distance) core-core interaction energy is attractive while the longer range interaction is repulsive. The former effect would tend to increase h_c while the latter would tend to increase λ_m with the consequences as discussed in the preceding paragraph.

Also analogously, the presence of a Peierls barrier or of frictional forces on dislocations arising from extrinsic defects in the crystals would contribute to the force balance or dislocations at local equilibrium. The effect again would be to tend to increase both h_c and λ_m.

V. Dislocation Spreading

In the growth of a multilayer structure, threading dislocations can grow into a growing layer and penetrate through it. Of interest is the critical condition for spreading of the dislocation into a dipole coating the interface, Fig. 10. As a specific example, we again consider a diamond cubic (0 0 1) interface and a glide dislocation on the system $\frac{1}{2}[0\ 1\ 1](1\ 1\ 1)$. The Peach-Koehler force on the segment D in Fig. 10b arising from the coherency stress is conveniently determined in coordinates fixed on the glide system, $i' = [0\ 1\ 1]/\sqrt{2}$, $j' = [1\ 1\ 1]\sqrt{3}$, and $k' = [2\ 1\ 1]/\sqrt{6}$. In these coordinates, $b = b, 0, 0$; $\sigma'_{12} = (\sqrt{6}/6)\sigma_{11}$ and the dislocation sense vector ξ is parallel to k'. The force on the dislocation is then the product of the Peach-Koehler force per unit length, $\sigma'_{12}b$, and

the segment length $h' = (\sqrt{6}/2)h$. For the single layer case treated in the earlier section where $c_A = c_0$, the resulting force is

$$F_{PK} = \mu b \kappa c_0 h \qquad (29)$$

For the multilayer case where $c_A = c_H = c_0/2$ the force is

$$F_{PK} = \mu b \kappa c_0 h/2 \qquad (30)$$

For the configuration of Fig. 10, the line tension force on the segment D, arising from the two dislocations on the interface, is

$$F_T = 2 \cdot \frac{\mu b^2 \kappa}{4\pi} \ln \left(\frac{h'}{r_0} \right) = \frac{\mu b^2 \kappa}{2\pi} \ln \left(\frac{3\sqrt{6}h}{2b} \right) \qquad (31)$$

with the κ term arising from end constraints as discussed previously. Equating Eqs. (29) and (31) for the single layer case, we find the equation for the critical thickness to be identical to the energy result of Eq. (28). For $h > h_c$, F_{PK} dominates and the dislocation spreads while it remains as a threading dislocation if $h < h_c$. The multilayer result follows from Eqs. (30) and (31).

Thus, the MB asymptotic energy result gives the exact answer for the critical thickness for spreading of an isolated threading dislocation in the simple line tension approximation. In the line tension approximation, the anisotropic elastic result is also simple enough in many cases to give an analytical result [26] for h_c. For closely spaced threading dislocations, dislocation interactions would enter the force balance and the result would differ: the effect would appear roughly when the spacing between threading dislocations became less than h or λ_c.

For a threading dislocation reaching the free surface of a single layer on a thick substrate, Fig. 11, F_{PK} is still given by Eq. (29). However, there is now only one

dislocation producing a line tension force and the outer cutoff radius h' is replaced by the image dislocation spacing $2h$. Thus

$$F_T = \frac{\mu b^2 \kappa}{4\pi} \ln\left(\frac{2h}{r_0}\right) = \frac{\mu b^2 \kappa}{4\pi} \ln\left(\frac{6h}{b}\right) \tag{32}$$

Thus, in this case the critical thickness is given by

$$h_c/\ln(6h_c/b) = b/4\pi\epsilon_0 \tag{33}$$

Comparing Eqs. (28) and (33), we see that spreading is easier for the layer at a free surface. This is of interest in practical applications in that it implies that substrate dislocations are more easily removed by sweeping for the first layer, favoring the growth of a perfect multilayer structure.

The line tension approximation is not exact because there is an interaction energy between segments D and E, for example, that comes into play in the initiation of spreading. The interaction energy can be determined for either internal layers or the free surface layer by standard methods [18] for dislocation segment interactions. However, at present, this extension of the force balance model has not been performed. Also, of course, forces associated with the Peierls barrier or with frictional effects arising from extrinsic defects, if present, would modify the force balance and the corresponding critical thickness. If these latter forces are large, dynamical effects could become important as discussed in the next section.

VI. Dislocation Injection

1. Nucleation

For a perfect layer under conditions where $h > h_c$, there are possible kinematic constraints to the generation of dislocations. The first of these is dislocation nucleation. People and Bean [36] draw attention to the possibility of such a constraint for strained layers. Their result led to an expression of the form $h_c/\ln(\text{const}\cdot h_c) \propto (1/\epsilon_o^2)$ in contrast to Eq. (17) or its later versions. However, in their expression for the energy of dislocation they introduced a typical (constant) dislocation spacing of 4 or 5 lattice spacings [36]. Instead, it is more appropriate to introduce the spacing $\lambda \propto (1/\epsilon_o)$, and then they would have recovered an expression for h_c analogous to the MB result.

Nucleation of perfect dislocations loops in the bulk has been considered by Frank [37] and that of perfect and partial dislocation half-loops at free surfaces by Hirth [38]. Matthews [7] presented a partial version of the latter work, deriving an expression for the critical-sized loop and discussing it but not considering the nucleation rate in detail. Maree et al. [39] have also reconsidered the problem, including the possibility of surface nucleation, but also focus on the loop energy and not the rate equations. Finally, nucleation at the edge of a multilayer [10] has been considered in the near-surface limit where σ_{12}, Eq. (6), is constant.

Here, we apply the earlier treatment [38] to nucleation in the interior of a strained layer where the complications of varying coherency stress fields are absent. The results qualitatively resemble those for the surface case. We first consider the nucleation of a loop of combined $\frac{1}{2}[0\ 1\ 1](1\ 1\ 1)$ dislocation within a strained multilayer for the example of a diamond cubic $(0\ 0\ 1)$ interface, Fig. 12a. The energy of the loop of radius r is [11, 12].

$$W_1 = \frac{\mu b^2 r}{4}\frac{(2-\nu)}{(1-\nu)}\left[\ln\frac{8r}{r_o} - 2\right] = \frac{\mu b^2 r}{4}\frac{(2-\nu)}{(1-\nu)}\ln\left(\frac{24r}{be^2}\right) \tag{31}$$

The coherency strain energy released is given in terms of the resolved coherency stress $\sigma'_{12} = (\sqrt{6}/6)\sigma_{11}$ of Fig. 10b by

$$W_2 = -\sigma'_{12}b\pi r^2 = -(\sqrt{6}/6)\pi b\mu\kappa\epsilon_o r^2 \tag{35}$$

Maximizing the total energy $W_1 + W_2$ with respect to r, we find for the radius of the critical sized loop that can grow with a decrease in energy

$$r_c = \frac{\sqrt{6}b}{8\pi\epsilon_0}\frac{(2-\nu)}{(1+\nu)}\ln\left(\frac{2.1r_c}{b_c}\right) \tag{36}$$

and for the energy of the critical sized loop

$$W_c = \frac{\mu b^2 r_c}{8}\frac{(2-\nu)}{(1-\nu)}\ln\left(\frac{2.1r_c}{b_c^3}\right) \tag{37}$$

Following standard rate theory [43], the concentration of critical nuclei is

$$n^* = n_1 exp(-W_c/kT) \tag{38}$$

where $n_1 \cong 1/b^3$ is the concentration of single atoms per unit volume. The nucleation rate is given by [44]

$$J = Z\omega n^* \tag{39}$$

where $Z \cong 0.1$ is the Zeldovich factor and

$$\omega = (8\pi r_c/b)\nu_D \tag{40}$$

is the frequency factor, with ν_D the Debye frequency. Typically, $\omega \cong 10^{15}s^{-1}$ and $n_1 \cong 10^{22}cc^{-1}$ so Eq. (39) becomes

$$J = 10^{36}exp(-W_c/kT)cc^{-1}s^{-1} \tag{41}$$

For nucleation to be important J must equal a critical observable value J^*. Here, unlike the earlier treatments, the critical value is one in a thin multilayer in a few thousand seconds or

one in a region viewed in an electron microscope: either estimate gives $J^* \simeq 10^{20} cc^{-1} s^{-1}$. The condition that must be satisfied for nucleation to be observable is given by Eq. (41) with J^* substituted for J or

$$W_c \leq 37kT \tag{42}$$

For the nucleation of a partial dislocation, there is an added energy term associated with the creation of a circle of stacking fault,

$$W_3 = \gamma \pi r^2 \tag{43}$$

where γ is the stacking fault energy. Since Eq. (35) and (43) have the same functional dependence on r, the same procedure as those for the perfect dislocation can be used if one replaces c_s in Eq. (35) by $c_n' = c_s - \gamma/\mu b$. Unless γ is quite large, partial dislocation nucleation is favored relative to the perfect dislocation case because b in Eq. (37) is smaller by a factor of 0.577.

To test the above cases, we use the same parameters as in a previous example [45], typical of compound semiconductors, $\mu = 48GPa, \nu = 0.23$ and $b = 0.100nm$ for the perfect dislocation; with $\gamma = 40 mJ/m^2$. The solution of the transcendental Eq. (42) then gives the results in Table I for r_c and c_s, the critical misfit to give appreciable nucleation. As seen in the table, nucleation is essentially impossible at room temperature, because the theoretical strength of a perfect crystal [46]($\sim \mu/10$) would be reached before the critical nucleation strength. Even at 1000 K nucleation is only a marginal possibility.

Another possibility, treated earlier [38] but neglected in later analyses [7, 39], is nucleation at the free surface of a growing layer. There are two possibilities: nucleation at a singular surface, Fig. 12b, and nucleation at a vicinal surface, Fig. 12c. For the

vicinal surface one must supply the added energy of the surface ledge

$$W_1 = 2r\epsilon \tag{44}$$

where ϵ is the ledge energy per unit length. However, only a portion of a loop need be created because of surface image interactions. The energies, $W_1, W_2,$ and W_3 are all reduced by a factor that is a function of the contact angle, θ, in turn determined by a balance of dislocation line tension and the surface ledge tension, approximately given by

$$\frac{\mu b^2}{2}\cos\theta = \epsilon \tag{45}$$

For calculation purposes, the typical value of ϵ is taken as 0.69 nJ/m.

The result for the vicinal surface is similar, but the ledge is removed during nucleation, so W_1, of the same form as in Eq. (45), becomes negative. The cumbersome equations for this case are given elsewhere [38, 17] and are not repeated here, but the results are also shown in Table 1. Even in this more favorable case, nucleation is a likely process only for large misfits in the vicinal surface-partial-dislocation case. The other possibility of nucleation at an edge [10], Fig. 12d, gives results quite close to those for the singular surface. A final possibility, discussed by Matthews [7], occurs if $r_c > h'$, whereupon the critical configuration becomes that of Fig. 12e. The exact energy is available for this configuration [48]. However, the values of r_c in Table 1 are so small that this case seems unlikely so that we do not proceed with the analysis.

In summary of the nucleation results, dislocation nucleation for the semiconductor case is only likely for the partial dislocation - vicinal surface case and otherwise should not occur. The physical basis for this result in comparison with the easier nucleation results for metal is the typical larger magnitude of b for the semiconductor case, a factor

that strongly influences W_2 in Eq. (37). This difficulty in nucleation is a favorable factor in the stability of strained layer structured. Yet, the presence of dislocations, presumably injected by nucleation, is often observed [7, 49]. We can only conclude that nucleation is augmented by extrinsic stresses, externally imposed or associated with defects. The resolved shear stress components τ of such defects would modify Eq. (35) to

$$W_2' = -(\sigma_{12}' + \tau)4\pi r^2 \qquad (46)$$

With stresses τ present, the critical misfit would be changed from the values in Table 1 to values c_n' given by

$$c_n' = c_n - \sqrt{6}\tau/\mu\kappa \qquad (47)$$

If τ were of the order of the coherency stresses, as would be the case for microcracks, surface asperities, disclinations [11], or inclusions, the critical misfit could be appreciably reduced.

2. Multiplication

The nucleation rate is so dependent on c_0 and τ that it changes from a negligible value to an enormous value over a small range of stress. Contrariwise, the rate of dislocation multiplication and the dislocation velocity are much less stress dependent. For example, under phonon or electron damping control the dislocation velocity is linear in stress. To take account of this weaker strain dependence, Dodson and Tsao [50] have introduced the useful concept of excess stress to drive the dislocation multiplication processes. If the critical coherency stress for thermodynamic stability is σ_0 when c_0 satisfies Eq. (18), for example, the driving force for dislocation motion would be zero when σ equalled σ_0. If

the coherency stress σ were greater than σ_o, the term σ_o would act as a drag stress and the driving force would be related to some function of the excess stress $(\sigma - \sigma_o)$.

In general, the critical stress for nucleation greatly exceeds σ_o so that if nucleation is the constraint to dislocation injection, multiplication and motion will be rapid and the excess stress considerations will be unimportant in controlling dislocation generation. If a few dislocations are already present in the structure, however, multiplication can occur at stresses well below those required for nucleation. Under these conditions, the concept of a critical excess stress to produce an appreciable rate of dislocation generation becomes important. Dodson and Tsao have shown that this leads to a kinetically controlled, required, critical thickness h_k greater than the thermodynamic value as schematically illustrated in Fig. 13. The excess stress monotonically increases with increasing strain rate or dislocation generation rate. Hence h_k is a kinetic factor such that h_K for a given dislocation content decreases with increasing time period of consideration.

VII. Summary

Equilibrium arrays of misfit dislocations in layer structures are treated in the Volterra dislocation model. Long range interactions lead to deviations from the early dipole results for the equilibrium critical layer thickness and have the important consequence that asymmetrical arrays are favored for the regime of a small ratio of layer thickness to misfit dislocation spacing. Once the critical thickness is reached, the energy decreases monotonically with dislocation spacing until the equilibrium spacing is reached so that there is no thermodynamic constraint to achieving the equilibrium array. Several cases of dislocation spreading are considered.

The nucleation of dislocations is considered. Nucleation in the absence of extrinsic forces or defects is found to be unlikely except for the case of a partial dislocation nucleating at a vicinal surface in a material with low stacking fault energy. The nucleation constraint leads to a lower bound in misfit below which nucleation is unlikely, a factor that should be of practical importance in the use of strained layer structure in applications. The interrelation of nucleation and the overstress needed for dislocation generation is considered.

Acknowledgement

References

1. A. Guinier. Compt. Rend., 206, 1641 (1938).
2. G.D. Preston, Proc. Roy. Soc. (London), A 167, 526 (1938).
3. F.C. Frank and J.H. van der Merwe, Proc. Roy. Soc. (London), A 198, 216 (1949).
4. A.J. Goss, K.E. Benson and W.G. Pfann, Acta Metall., 4, 332 (1956).
5. P. Delavignette, J. Tournier, and S. Amelinckx, Philos. Mag., 6, 1419 (1961).
6. J.W. Matthews, Philos. Mag., 6, 1317 (1961).
7. J.W. Matthews, in "Dislocations in Solids," (F.R. N. Nabarro, ed.), Vol. 2, p. 461. North Holland, Amsterdam, 1979.
8. G.C. Osbourn, J. Vac. Sci., B1, 379 (1983).
9. G.C. Osbourn, Phys. Rev., B27, 5126 (1983).
10. U. Gradmann, Ann. Phys. Lpz., 13, 213 (1964).
11. W. Bollman, "Crystal Defects and Crystal Interfaces," Springer-Verlag, Berlin, 1970.
12. J.P. Hirth and R.W. Balluffi, Acta Metall., 21, 929 (1973).
13. R.W. Balluffi and G.B. Olson, Metall. Trans., 16A, 529 (1985).
14. R.C. Pond, in "Dislocations in Solids," (F.R.N. Nabarro, ed.), Vol. 8, North-Holland, Amsterdam, in press.
15. J.H. van der Merwe and W.A. Jesser, J. Appl. Phys. 64, 4968 (1988).
16. J.H. van der Merwe, Proc. Phys. Soc. (London), A63, 616 (1950).

17. J.W. Matthews and A.E. Blakeslee, J. Cryst. Growth, 27, 118 (1974); 29, 273 (1975); 32, 265 (1976).
18. J.P. Hirth and J. Lothe, "Theory of Dislocations," 2nd edition, Chapter 8, Wiley, New York, 1982.
19. J.H. van der Merwe, J. Appl. Phys., 34, 123 (1963).
20. W.A. Jesser and J.H. van der Merwe, J. Appl. Phys., 63, 1509, 1928 (1988).
21. J.E. Sinclair, P.C. Gehlen, R.G. Hoagland and J.P. Hirth, J. Appl. Phys., 49, 3890 (1978).
22. B.W. Dodson, Phys. Rev., B35, 2795, 5558 (1987).
23. J.R. Willis, Philos. Mag., 21, 931 (1970).
24. A.N. Stroh, Philos. Mag., 3, 625 (1958).
25. D.M. Barnett and J. Lothe, Phys. Norvegica, 7, 13 (1973).
26. J.P. Hirth, S. Afr. J. Phys., 9, 72 (1986).
27. S.V. Kamat, J.P. Hirth and B. Carnahan, Scripta Metall., 21, 1587 (1987).
28. W. Mader, Mat. Res. Soc. Proc., 82, 403 (1987).
29. J.P. Hirth, in "Single Crystal Films," (M. Francombe and H. Sato, eds.), p. 173. Pergamon, Oxford, 1964.
30. C.A.B. Ball and C. Laird, Thin Solid Films, 41, 307 (1977).
31. F.J.J. van Loo, B. Pieraggi and R.A. Rapp, Acta Metall., in press.
32. M.D. Drory, M.D. Thouless and A.G. Evans, Acta Metall., 36, 2019 (1988).
33. K. Shinohara and J.P. Hirth, Philos. Mag., 27, 883 (1973).
34. J.P. Hirth and J. Lothe, "Theory of Dislocations," 2nd edition, Chapter 19, Wiley, New York, 1982.
35. X. Feng, PhD Dissertation, Washington State University, Pullman, WA.
36. R. People and J.C. Bean, Appl. Phys. Lett., 47, 322 (1985); 49, 229 (1986).
37. F.C. Frank, in "Symposium on Plastic Deformation of Crystalline Solids," p. 89, Carnegie Inst. of Tech., Pittsburgh, 1950.
38. J.P. Hirth, in "The Relation Between the Structure and Mechanical Properties of Alloys," p. 217, H.M. Stationery Off., London, 1963.
39. P.M.J. Maree, J.C. Barbour, J.F. van der Veen, K.L. Kavanagh, C.W.T. Bulle-Lieuwma and M.P.A. Viegers, J. Appl. Phys., 62, 4413 (1987).
40. J.P. Hirth and A.G. Evans, J. Appl. Phys., 60, 2372 (1986).
41. F.R.N. Nabarro, Adv. Phys., 1, 332 (1952).
42. J.P. Hirth and J. Lothe, "Theory of Dislocations," 2nd edition, p. 169, Wiley, New York, 1982.
43. J. Feder, K.C. Russell, J. Lothe and G.M. Pound, Adv. Phys., 15, 111 (1966).
44. T.L. Davis and J.P. Hirth, J. Appl. Phys. 37, 2112 (1966).
45. H. Ehrenreich and J.P. Hirth, Appl. Phys. Lett., 46, 668 (1985).
46. A. Kelly, "Strong Solids," 2nd edition, Clarendon, Oxford, 1973.
47. S.V. Kamat and J.P. Hirth, manuscript in preparation.
48. J.P. Hirth and J. Lothe, "Theory of Dislocations," 2nd edition, p. 113, Wiley, New York, 1982.
49. Z.J. Radzinski, B.L. Jiang, G.A. Rozgonyi, T.P. Humphreys, N. Hamaguchi and S.M. Bedair, J. Appl. Phys., 61, 2328 (1988).
50. B.W. Dodson and J.Y. Tsao, Appl. Phys. Lett., 51, 1325 (1987).

TABLE I. Values of r_c/b and c_n^* for several cases
of dislocation nucleation

| | Perfect Disl. | | | | Partial Disl. | | | |
| | 293 K | | 1000 K | | 293 K | | 1000 K | |
	r_c/b	c_n^*	r_c/b	c_n^*	r_c/b	c_n^*	r_c/b	c_n^*
Internal loop	0.98	0.31	1.31	0.26	1.51	0.13	2.61	0.30
Singular surface	0.73	0.13	1.18	0.31	1.21	0.57	2.55	0.33
Vicinal surface	2.85	0.12	3.60	.0.11	7.17	0.11	11.0	0.087

Figure Captions

Figure 1. Representation of (a), coherent interface, (b), relaxed crystals with misfit dislocations, (c), misorientation dislocations, and (d), combined dislocations.

Figure 2. Projection of simple cubic lattices along $[0\ 0\ \bar{1}]$ parallel to a $(3\ 1\ 0)$ twin boundary. Grid is the DSC lattice and balls are atom positions. Different atom configurations across twin plane on two sides of II) indicate different fault energies.

Figure 3. Ledge translational dislocations at an interface ledge with Burgers vector equal to the difference in interplanar spacing Δd at the interface.

Figure 4. (a),alternating stresses in a strained layer structure in a fcc crystal, and (b), projection of a single A layer.

Figure 5. Surface tractions at free edge of specimen and resulting shear stresses σ_{13}.

Figure 6. Creation of set I and set II dislocations illustrated for one dipole: creation at I and separation via 2 to equilibrium position 3.

Figure 7. Equilibrium arrays for h slightly greater than h_c: (a), $h_c < \lambda$, (b), $h_c > \lambda$.

Figure 8. Plot of h_1/λ versus b/λ for the single dipole case of Eq. (16) and for the multiple dipole case, Eqs. (21) and (23). Values of X_c are noted for the multiple dipole case.

Figure 9. Combined dislocations formed by glide on $[0\ 1\ 1](1\ 1\ 1)$: (a), single dipole, (b), multiple dipole.

Figure 10. View parallel to the $(0\ 0\ 1)$ interface of the spreading of a $\frac{1}{2}\ [0\ 1\ 1]$ dislocation on $(1\ 1\ \bar{1})$: (a), view along the direction $[\bar{1}\ 1\ 0]$ of spreading and parallel to the glide plane, (b) view along $[\bar{1}\ \bar{1}\ 0]$, perpendicular to b, to ξ, and to the direction of spreading.

Figure 11. Free surface version of Fig. 10. Image dislocation shown as dashed line.

Figure 12. Critical-sized glide dislocation nuclei for several cases with the Burgers vector in

the plane of the loop: (a), nucleation within bulk A layer, (b), nucleation at a singular

surface, requiring ledge creation, (c), nucleation at a vicinal surface, removing a ledge,

(d), nucleation in the interface plane at a crystal edge and, (e), nucleation within an

A layer for the case that $r_e > h'$.

a.

b.

c.

d.

a.

b.

a.

B

h_A

A

λ

B

b.

B

$\lambda/2$

h_A

A

λ

B

a.

$\frac{1}{2}[0\bar{1}1]$ (001)

b.

A   B

a.

b.

a.

$\frac{1}{2}[0\bar{1}1]$ (001)

b.

A   B

a.

B

A

B

h'

h

b.

E

D

F_{PK}

$\vec{\xi}$

b

a.

b.

a.

b.   SURFACE
     LEDGE
     θ
     r

c.   SURFACE
     LEDGE
     r
     θ

d.   r

e.   h   r
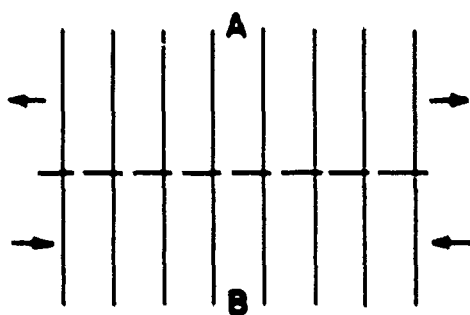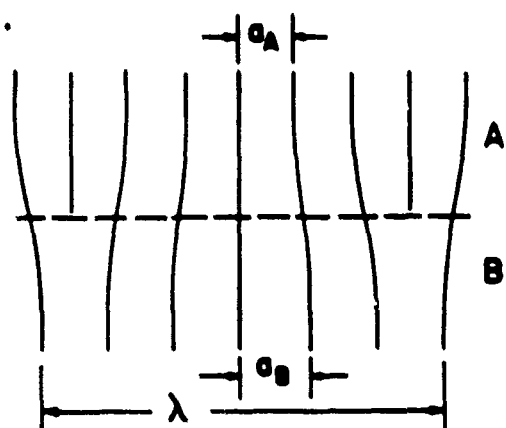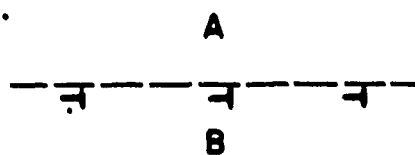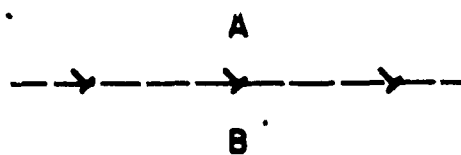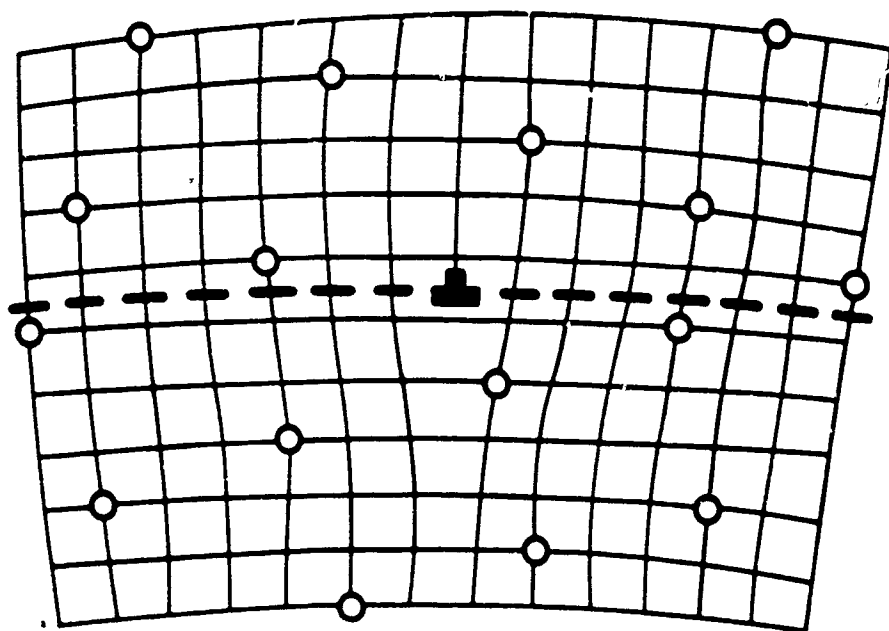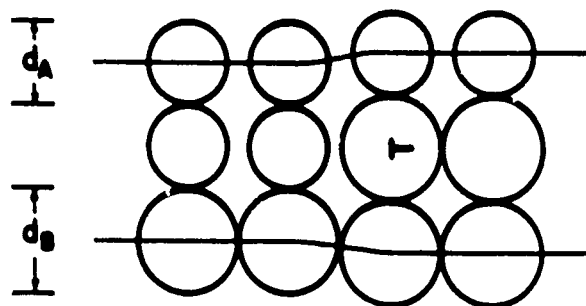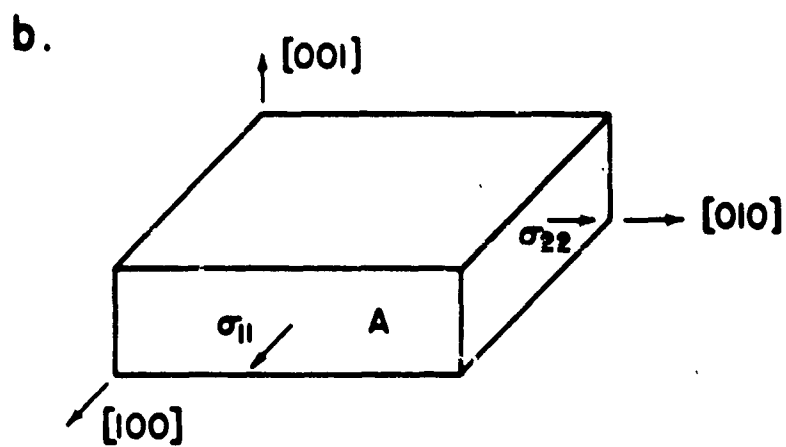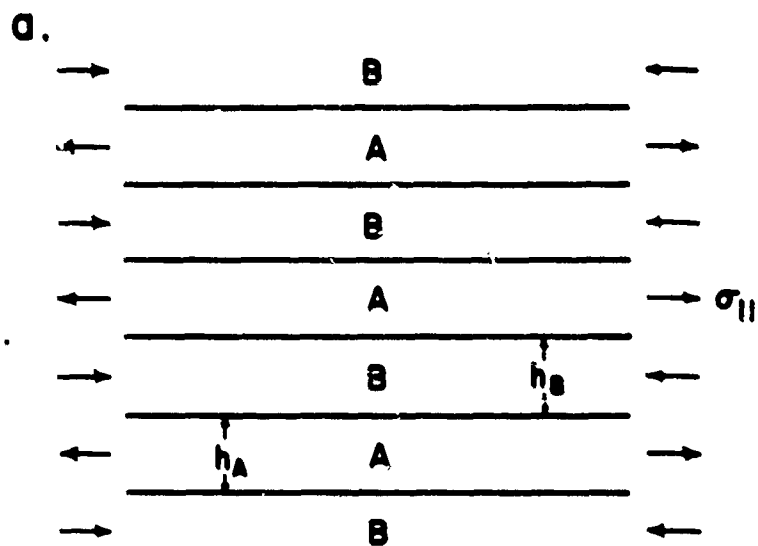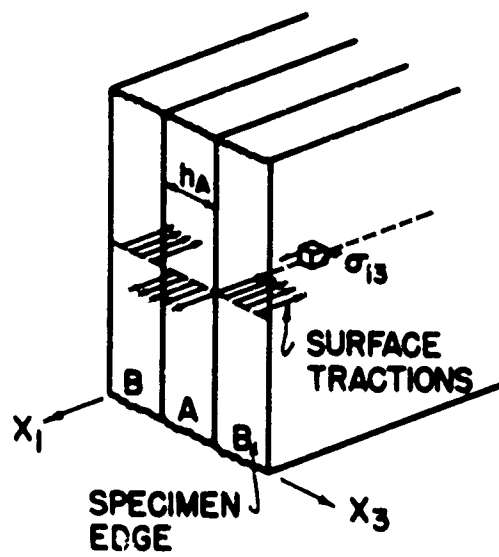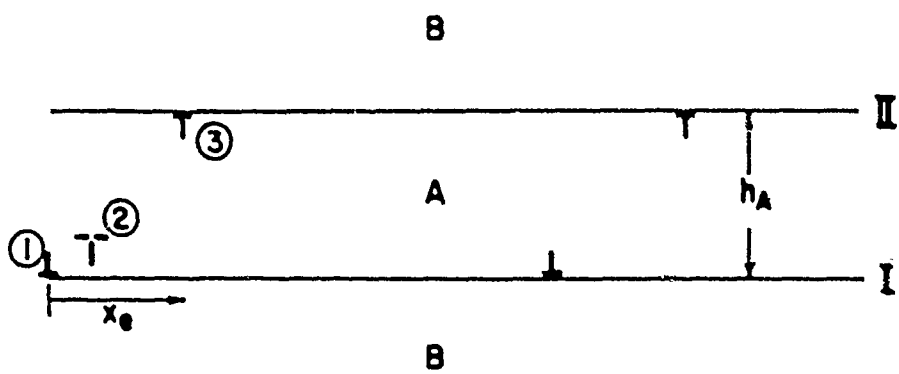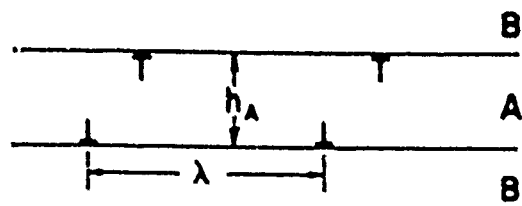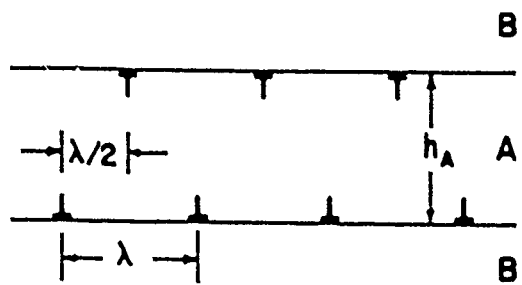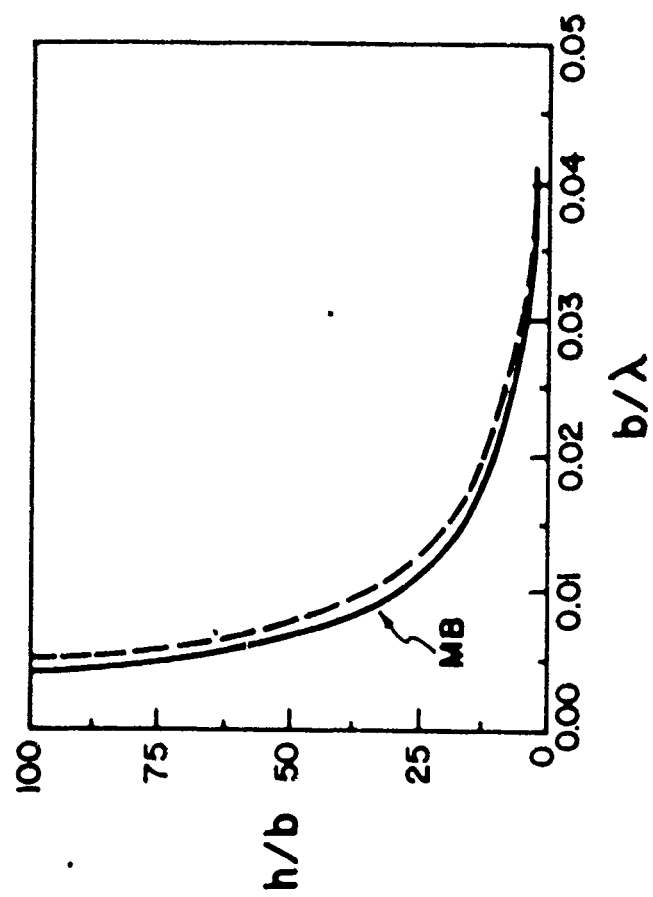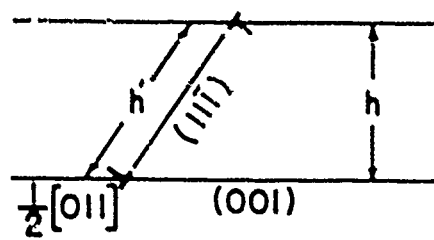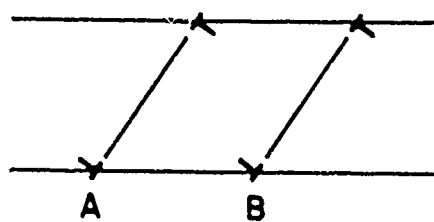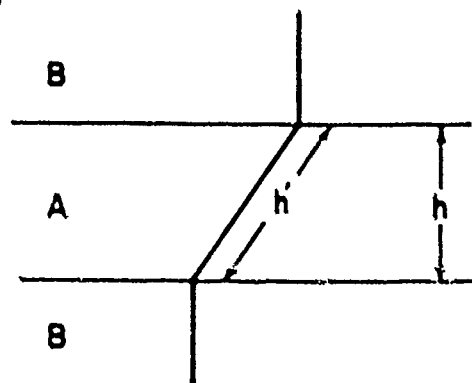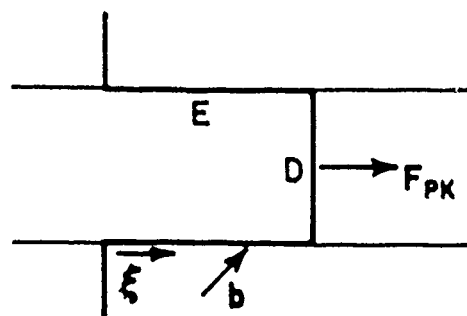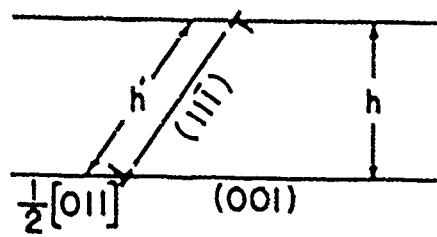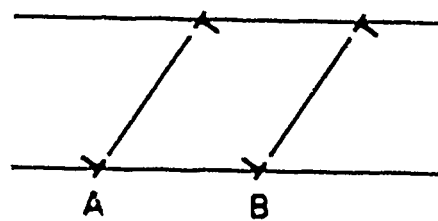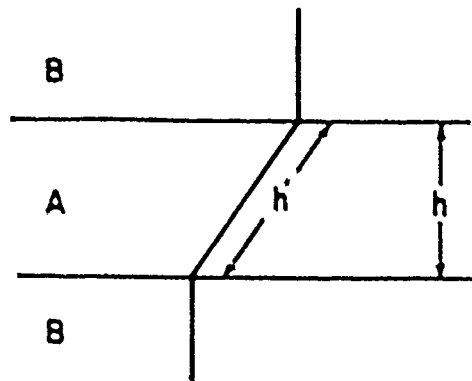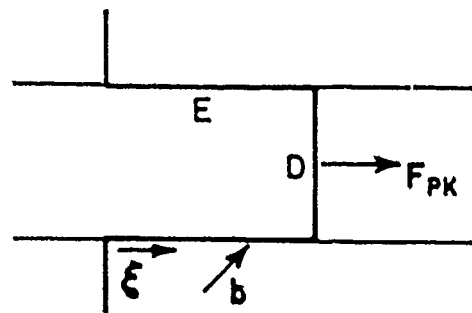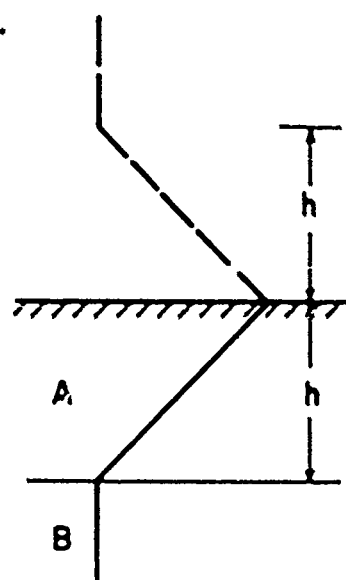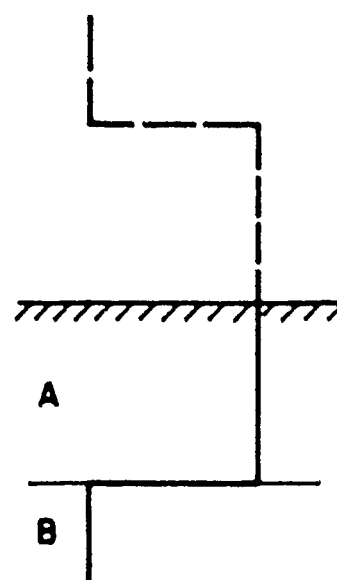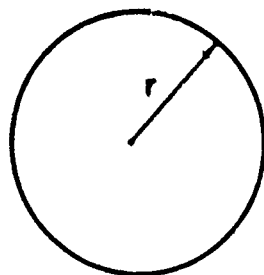
DISLOCATION INJECTION IN STRAINED MULTILAYER STRUCTURES

S.V. Kamat and J.P. Hirth
Mechanical and Materials Engineering Dept.
Washington State University
Pullman, WA 99164-2920

.

(PACS 81-98)

ABSTRACT

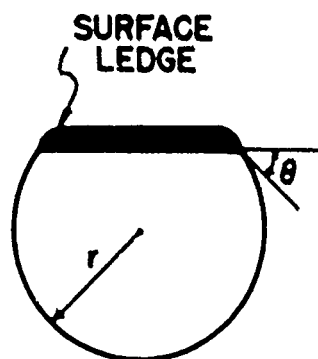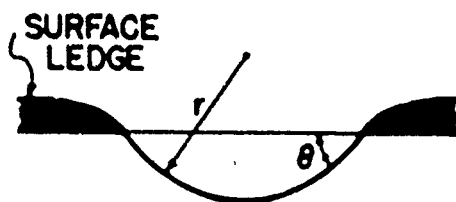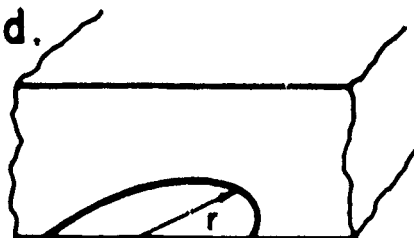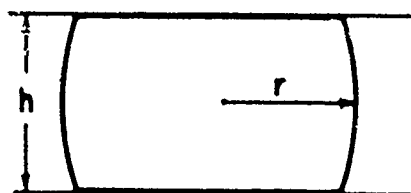Nucleation of dislocations within a strained multilayer and at a singular or vicinal multilayer surface is considered for typical compound semiconductor and metal cases. The results indicate that strained multilayer structures are stable against dislocation injection, even when the layer thickness exceeds the thermodynamic critical thickness for misfit dislocation stability, in the absence of defects or large external stresses. An exact energy calculation for spreading of a threading dislocation in GaAs layer is also presented. The results show that in the absence of any jog formation, there is no barrier to spreading at or above the thermodynamic, Matthews-Blakeslee critical layer thickness. Moreover, reversible spreading to an equilibrium standoff distance occurs for thicknesses smaller than the critical thermodynamic value an effect that can lead to dissipative reversed dislocation motion under alternating driving forces.

I. INTRODUCTION

In recent years there has been renewed interest in the study of strained multilayer structures with alternating composition or phases because of their potential for electronic applications.[1] Many of these applications require the multilayer structures to be dislocation free.[2] However, the misfit caused by the lattice mismatch between adjacent layers in a multilayer structure enhances the likelihood of dislocation formation. Observations indicate that the accommodation of misfit between layers in a multilayer structure is purely by elastic strains below a critical thickness of the layers whereas it can be shared between dislocations and elastic strains above this critical thickness.[3-5] However, the fraction of misfit accommodated by dislocations is often found to be less than that expected on the basis of purely thermodynamic considerations, probably because of kinematic constraints to dislocation formation.

Nucleation is one of the possible kinematic constraints to the generation of dislocations in strained multilayer structures. Nucleation of perfect dislocation loops in the bulk has been previously discussed by Frank[6] and that of perfect and partial dislocation half-loops at a free surface by Hirth.[7] Matthews[8] presented a partial version of the latter work, deriving an expression for the critical-sized loop and discussing it but not considering the nucleation rate in detail. Marée et al[9] have also reconsidered the problem, including the possibility of surface nucleation, but they also focus on the loop energy and not the rate equations. Finally nucleation at the edge of a multilayer has been treated [10] in the region of varying shear stress in the approximation that the shear stress has a constant value equal to the limiting value at the free lateral surface.

In this treatment we consider the nucleation of perfect and partial dislocation loops both in the interior and on two types of free surface parallel to the multilayer interfaces: on a singular surface, and on a vicinal surface, with emphasis placed on the rate equations. The example selected is that of a diamond cubic, or metallic, (001) interface as shown in Fig. 1. The loops are assumed to be formed in a region where the shear stress associated with the free edge surface boundary conditions are negligibly small and hence the complications of varying coherency stress fields are absent. The earlier treatments [6-9] are extended by including the influence of the line energy associated with the surface ledges that participate in the surface nucleation events.

We also consider the spreading of a threading dislocation in an exact-energy model for the bowing dislocation, explicitly including the interaction energy of the moving dislocation segments with the remainder of the threading dislocation.

II. NUCLEATION FORMULATION

The procedure in each case involves calculating the total energy associated with the nucleation of a dislocation loop which is the algebraic sum of the self-energy of the loop, the coherency strain energy released because of the formation of the loop and any energy required for creating a new surface step or a stacking fault. Maximizing this total energy with respect to the loop radius r, one can find the radius (r_c) and energy (W_c) for a critical-sized loop that can grow with a decrease in free energy.

In accord with standard rate theory[11], the concentration of critical nuclei (n^*) is given by

$$n^* = n_1 exp(-W_c/kT) \tag{1}$$

where $n_1 \cong 1/b^3$ is the concentration of single atoms per unit volume, k is the Boltzmann constant and T is the absolute temperature. The nucleation rate is given by

$$J = Z\omega n^* \qquad (2)$$

where $Z \cong 0.1$ is the Zeldovich factor and $\omega = (8\pi r_c/b)\nu_b$ is the frequency factor with ν_b the Debye frequency. Typically $\omega \approx 10^{13} s^{-1}$ and $n_1 \cong 10^{22} cc^{-1}$ so Eq.(2) becomes

$$J = 10^{36} \exp{(-W_c/kT)} cc^{-1} s^{-1} \qquad (3)$$

For nucleation to be important, J must equal a critical observable value J^*. Here, unlike the earlier treatments,[6,7] we take the critical rate as one nucleation event in a thin multilayer in a few thousand seconds or one event in a region of 10^{-15} to $10^{-18} cc$, typical of a region viewed in an electron microscope: either estimate gives $J^* \cong 10^{20} cc^{-1} s^{-1}$. The condition that must be satisfied for nucleation to be observable is thus given by Eq.(3) with J^* substituted for J or

$$W_c \leq 37\, kT \qquad (4)$$

The solution of Eq.(4) gives the critical radius and misfit (c_o^*) needed for nucleation to be observable in strained multilayers. Let us now consider the individual cases to obtain the expressions for r_c and W_c.

A. Interior loop

1. Perfect Dislocation

We consider the case of a nucleation of a loop of a combined $\frac{1}{2}[011](11\bar{1})$ perfect dislocation, as shown in Fig. 2, in the approximation that the elastic constants of the two materials in the layer structures are the same. Also we consider the usually observable case[8] of a glide dislocation on the $(12\bar{1})$ plane inclined to the (001) interface.

The self-energy of the loop of radius r is given by

$$W_1 = \frac{\mu b^2 r}{4} \frac{(2-\nu)}{(1-\nu)} \ell n \left(\frac{24r}{be^2}\right) \qquad (5)$$

where μ is the shear modulus, ν is the Poisson's ratio, b is the magnitude of the Burgers vector and e is the Naperian logarithmic base. Here we have taken the core cutoff radius to be that characteristic[12] of diamond-cubic materials, $r_o = b/3$. The coherency strain energy released because of the formation of the loop is given by

$$W_2 = -\sigma_{12}' b\pi r^2 = \frac{-\sqrt{6}}{6}\pi b\mu\kappa c_o r^2 \qquad (6)$$

where $\kappa = \frac{(1+\nu)}{(1-\nu)}$. Here, see Fig. 2, $c_o = \Delta a/<a>$ is the misfit strain, with Δa the difference and $<a>$ the average of the lattice parameters of the two materials and σ_{12}' is the shear stress resolved on the glide plane and in the glide direction. We suppose the thicknesses of the multilayer components are equal. Other cases can be easily determined from the following results by the use of alternative forms relating c_o and Δa, see ref. (10).

Maximizing the total energy ($W_1 + W_2$) gives the radius of the critical-sized loop (r_c) as

$$r_c \cong \frac{3b}{4\sqrt{6}\pi c_o} \frac{(2-\nu)}{(1+\nu)} \ell n \left(\frac{24 r_c}{be}\right) \tag{7}$$

and the energy of the critical-sized loop (W_c) as

$$W_c = \frac{\mu b^2 r_c}{8} \frac{(2-\nu)}{(1-\nu)} \ell n \left(\frac{24 r_c}{bc^3}\right) \tag{8}$$

2. Partial Loop

In diamond cubic structures the perfect dislocation $\frac{1}{2}[011]$ can dissociate into partials $\frac{1}{6}[112]$ and $\frac{1}{6}[121]$. Hence, the nucleation of a partial loop is of interest, for which

$$W_1 = \frac{\mu b_p^2 r}{4} \frac{(2-\nu)}{(1-\nu)} \ell n \left(\frac{24 r}{b_p c^2}\right) \tag{9}$$

where b_p is the magnitude of the Burgers vector of the partial dislocation and is equal to $0.577b$ and,

$$W_2 = \frac{-\sqrt{2}}{6} \pi \mu b_p \kappa c_o' r^2 \tag{10}$$

where

$$c_o' = \left(c_o - \frac{3\sqrt{2}\gamma_{sf}}{\mu b_p \kappa}\right) \tag{11}$$

with γ_{sf} being the stacking fault energy and the term containing γ_{sf} being the modification in the energy associated with the area πr^2 of created stacking fault. Maximizing the total energy we find,

$$r_c = \frac{3b_p}{4\sqrt{2}\pi c_o'} \frac{(2-\nu)}{(1+\nu)} \ell n \left(\frac{24 r_c}{b_p c}\right) \tag{12}$$

$$\text{and } W_c = \frac{\mu b_p^2 r_c}{8} \frac{(2-\nu)}{(1-\nu)} \ell n \left(\frac{24 r_c}{b_p c}\right) \tag{13}$$

B. Singular Surface

1. Perfect Dislocation loop

Consider the nucleation of a loop of a perfect $\frac{1}{2}[011]$ ($11\bar{1}$) dislocation on a singular free surface of a multilayer as shown in Fig. 3. Because of the energy $c \cong \gamma b$ of the created surface ledge, where γ is the surface energy of the ledge face, the loop of radius of curvature r will meet the surface at a contact angle θ determined by a balance of ledge line tension c with the dislocation line tension:

$$c \cong \gamma b \cong \frac{1}{2}\mu b^2 \cos\theta \tag{14}$$

the energy for the formation of such a loop is

$$W_1 = \frac{\mu b^2 r}{4} \frac{(2-\nu)}{(1-\nu)} \frac{(\pi-\theta)}{\pi} \ln\left(\frac{24r}{bc^2}\right) \tag{15}$$

The coherency strain energy released is

$$W_2 = \frac{-\sqrt{6}}{6}\mu b \kappa c_o r^2 [(\pi-\theta) + \cos\theta\sin\theta] \tag{16}$$

There is also an additional energy term associated with the formation of the surface ledge that is given by

$$W_3 = 2r\gamma b\sin\theta \tag{17}$$

Maximizing the total energy ($W_1 + W_2 + W_3$), one finds,

$$r_c = \frac{\sqrt{6}}{8c_o} b \frac{(2-\nu)}{(1+\nu)} \frac{(\pi-\theta)}{\pi[(\pi-\theta)+\cos\theta\sin\theta]} \ln\left(\frac{24}{b}e^{\lambda-2}r_c\right) \tag{18}$$

where $\lambda = \left[\frac{8\gamma\sin\theta}{\mu b}\frac{(1-\nu)}{(2-\nu)}\frac{\pi}{(\pi-\theta)}\right] + 1$

and $W_c = \frac{\mu b^2}{8}\frac{(2-\nu)}{(1-\nu)}\frac{(\pi-\theta)}{\pi}r_c \ln\left(\frac{24}{b}e^{\lambda-4}r_c\right) \tag{19}$

2. Partial Dislocation Loop

For the case of a partial dislocation loop nucleating on a singular surface, the energy for the creation of the loop is given by

$$W_1 = \frac{\mu b_p^2 r}{4}\frac{(2-\nu)}{(1-\nu)}\frac{(\pi-\theta)}{\pi}\ln\left(\frac{24r}{b_p c^2}\right) \tag{20}$$

The coherency strain energy released is given by

$$W_2 = \frac{-\sqrt{2}}{6}\mu b_p \kappa c_o' r^2[(\pi-\theta)+\cos\theta\sin\theta] \tag{21}$$

where c_o' is given by equation (11). The energy for the creation of the surface ledge (W_3) is

$$W_3 = 2r\gamma b_p\sin\theta \tag{22}$$

Maximizing the total energy we find

$$r_c = \frac{3\sqrt{2}}{8}\frac{b_p}{c_o'}\frac{(2-\nu)}{(1+\nu)}\frac{(\pi-\theta)}{\pi[(\pi-\theta)+\cos\theta\sin\theta]}\ln\left(\frac{24e^{\lambda-2}r_c}{b_p}\right) \tag{23}$$

where $\lambda = \left[\frac{8\gamma\sin\theta}{\mu b_p^2}\frac{(1-\nu)}{(2-\nu)}\frac{\pi}{(\pi-\theta)}\right]+1$

and $W_c = \frac{\mu b_p^2}{8}\frac{(2-\nu)}{(1-\nu)}\frac{(\pi-\theta)}{\pi}r_c\ln\left(\frac{24e^{\lambda-4}r_c}{b_p}\right) \tag{24}$

C. Vicinal Surface

1. Perfect Dislocation Loop

Consider the nucleation of a loop of a perfect $\frac{1}{2}$ [011] (11$\bar{1}$) dislocation on a vicinal surface of a multilayer as shown in Fig. 4. The value of θ is given by equation (14) for this case also, resulting from the line tension balance for the case of Fig. 3. The energy for creation of such a loop is

$$W_1 = \frac{\mu b^2 r}{4}\frac{(2-\nu)}{(1-\nu)}\frac{\theta}{\pi}\ell n\left(\frac{12R}{bc^2}\right) \tag{25}$$

Here the cutoff distance R is, instead of the loop diameter $2r$, the normal distance from the deepest part of the loop to the corresponding image dislocation and is given by

$$R = 1.634r\,(1 - \cos\theta) \tag{26}$$

The coherency strain energy released is

$$W_2 = \frac{-\sqrt{6}}{6}\mu b\kappa c_o r^2[\theta - \sin\theta\cos\theta] \tag{27}$$

The energy associated with forming the surface ledge is again given by Eq. (17).
Maximizing the total energy with respect to r, we determine

$$r_c = \frac{\sqrt{6}}{8c_o}b\frac{(2-\nu)}{(1+\nu)}\frac{\theta}{\pi}\ell n\left(\frac{19.60(1-\cos\theta)r_c}{b\,e^{\lambda'+2}}\right) \tag{28}$$

where $\lambda' = \left[\frac{8\gamma\sin\theta}{\mu b}\frac{(1-\nu)}{(2-\nu)}\frac{\pi}{\theta}\right] - 1$

$$\text{and } W_c = \frac{\mu b^2 r_c}{8}\left(\frac{2-\nu}{1-\nu}\right)\frac{\theta}{\pi}\ell n\left(\frac{19.60(1-\cos\theta)r_c}{b\,e^{\lambda'+1}}\right) \tag{29}$$

2. Partial Dislocation Loop

For a partial dislocation loop nucleating on a vicinal surface

$$W_1 = \frac{\mu b_p^2 r}{4}\frac{(2-\nu)}{(1-\nu)}\frac{\theta}{\pi}\ell n\left(\frac{12R}{b_p c^2}\right) \tag{30}$$

$$W_2 = \frac{-\sqrt{2}}{6}\mu b_p\kappa c_o' r^2[\theta - \sin\theta\cos\theta] \tag{31}$$

$$\text{and } W_3 = 2r\gamma b_p\sin\theta \tag{32}$$

Maximizing the total energy with respect r, we find

$$r_c = \frac{3\sqrt{2}b_p}{8c_o'}\frac{(2-\nu)}{(1+\nu)}\ell n\left(\frac{19.60(1-\cos\theta)r_c}{b_p e^{\lambda'+2}}\right) \tag{33}$$

where $\lambda' = \left[\frac{8\gamma\sin\theta}{\mu b_p}\frac{(1-\nu)}{(2-\nu)}\frac{\pi}{\theta}\right] - 1$

$$\text{and } W_c = \frac{\mu b_p^2 r_c}{8}\frac{(2-\nu)}{(1-\nu)}\frac{\theta}{\pi}\ell n\left(\frac{19.60(1-\cos\theta)r_c}{b_p e^{\lambda'+1}}\right) \tag{34}$$

III. THREADING DISLOCATION MOTION

We consider the same glide system as in Fig. 2, with the bowout configuration as shown in Fig. 5a and b. Also, we suppose that the threading dislocation is initially inclined on the glide system so as to be initially straight. This will give a lower bound to the blowout energy since added energy would be required to create a jog E, Fig. 5c, if the threading dislocation were initially normal to the interface. The total energy of this configuration is given by an alternate limit of the result given for a double kink in ref. (13).

$$W = \lim_{M \to \infty} \Delta W \tag{35}$$

$$\text{where } \Delta W = 2W_{int}(CA) - 2W_{int}(C'A) + W_{int}(D, E) \\ + 2W_s(D) \tag{36}$$

Here W_s is the self energy of segment D and $W_{int}(CA)$, for example, is the interaction energy between segments C and A. Exact values, within the isotropic elastic approximation, are available [13] for the segment interaction energies and self-energies. Taking the limit $M \to \infty$, one finds

$$W_1 = \frac{\mu b^2}{8\pi} \frac{(4 - 3\nu)}{(1 - \nu)} [(h'^2 + a^2)^{\frac{1}{2}} - h' - a + h' \ell n \left(\frac{2h'}{h' + (h'^2 + a^2)^{\frac{1}{2}}} \right)]$$

$$- \frac{\mu b^2}{8\pi} \frac{(4 - \nu)}{(1 - \nu)} [h' - (h'^2 + a^2)^{\frac{1}{2}} + a \ell n \left(\frac{a + (h'^2 + a^2)^{\frac{1}{2}}}{h'} \right)] + \frac{\mu b^2 a}{8\pi} \frac{(4 - \nu)}{(1 - \nu)} \ell n \left(\frac{a}{c\rho} \right) \tag{37}$$

where $\rho = \frac{c_a}{2} = \frac{b}{6}$ and $h' = \frac{\sqrt{6}}{2} h$. The coherency strain energy released because of the spreading is given by

$$W_2 = -\mu b \kappa \frac{c_o}{2} ha \tag{38}$$

In the limit $a \gg h$, the Matthews-Blakeslee result for the critical thickness h^*, above which spreading should occur in a simple line-tension model, is obtained by setting $W_1 = W_2$. The result is

$$h^* = \frac{b}{4\pi c_o} \frac{(4 - \nu)}{(1 + \nu)} \ell n \left(\frac{\alpha}{c\rho} \right) \tag{39}$$

IV. RESULTS AND DISCUSSION

A. Nucleation

The radius of the critical-sized loop (r_c) and the critical misfit (c_o^*) required for nucleation to be observable were calculated for all the above cases for a gallium arsenide (GaAs) and a silver (Ag) layer in the respective multilayer systems: GaAs-X and Ag-X. Here X would represent the respective semiconductor compound or metal for the second constituent of the multilayer. The properties used in the calculations for the GaAs layer were $\mu = 48 GPa$, $\nu = 0.23$, $b = 0.4 nm$, $\gamma = 1.725 \ kJ/m^2$ and $\gamma_{sf} = 40 \ mJ/m^2$ while those for the Ag layer were $\mu = 33.8 \ GPa$, $\nu = 0.354$, $b = 0.2889 nm$, $\gamma = 1.140 \ kJ/m^2$

and $\gamma_{sf} = 16mJ/m^2$. Also for the silver case the cutoff radius is that characteristic in a metal, $r_o = \frac{b}{4}$, so a factor 3 appears in the denominator of the logarithmic term for all W_i equations. The values of $\frac{b}{r}$ and c_o^* at 293K and 1000K are listed in Tables 1 and 2 for the GaAs layer and the Ag layer, respectively.

Tables 1 and 2 reveal that nucleation is more likely at 1000K than at 293K as indicated by the relatively lower misfits required at 1000K. When all the other factors are equal, nucleation on a vicinal surface is more favorable than nucleation of an interior loop or nucleation on a singular surface. Also, the critical misfits required for nucleation to be important are relatively lower for the Ag layer than for the GaAs layer. In most compound semiconductors the misfit is typically less than 0.01. A comparison with the critical misfits for the GaAs layer for any type of nucleation shows that the critical misfits are much larger than 0.01. This suggests that nucleation of dislocation loops is highly unlikely in compound semiconductors such as the GaAs-X system in the absence of some extrinsically imposed external stress or of a defect such as a microcrack. The present results closely parallel those for nucleation at a free edge surface[10] in showing that nucleation is highly unlikely unless defects or large external stresses are present. The above solutions are valid when the critical-sized loop lies entirely within a given layer, that is when $r_c < h'/2 < \sqrt{\frac{5}{4}}h$. When the radius exceeds $h'/2$, the critical configuration would be that in Fig. 6, as discussed by Matthews.[7] However, as indicated in Tables 1 and 2, the small values of r_c for all cases of nucleation are such that the condition $h'/2 < r_c$ is very unlikely. Were h to be so small that $h'/2 < r_c$, nonlinear elastic interactions between interfaces would become important along with other second-order effects not considered in the present treatment.

B. Threading Dislocation

The properties used in the calculations for spreading of a threading dislocation are those given above for a GaAs layer in a GaAs-X multilayer system. Plots of total energy (W) versus the dislocation spreading length (a) for several layer thicknesses h are shown in Figs. 7-9. For values of h below a critical value, the total energy W initially decreases with increasing a, goes through a minimum and then increases monotonically. Above a critical value, h^*, the total energy monotonically decreases with increasing a. The critical layer thickness was found to be 61.85 nm in agreement for the present geometry with the Matthews-Blakeslee limiting value[3-5] of h^*, equivalent to the infinite dipole result of Eq. (39).

Above the critical thickness h^*, the dislocation spreads spontaneously with a monotonic decrease in free energy, equivalent to the Matthews-Blakeslee result.[3-5] Below the critical thickness there is some spreading to an equilibrium position a^* given in Fig. 10. This spreading has an analog in the reversible bowing of a pinned dislocation under bowout stresses less than the critical value for breakway. Hence, analogous to the pinned dislocation case, there could be dissipative reversed dislocation motion for the multilayer analogous to the Granato-Lücke process for pinned dislocations.[14]

For the configuration of Fig. 5c, the W-a relation is equivalent to that in Figs. 7-9 for large a. However, because of the necessity of creating the jog, there will be an energy barrier of the order of a jog formation energy to spreading even when the thickness is equal to the Matthews-Blakeslee critical value.

V. SUMMARY

Nucleation of dislocations within a strained multilayer and at a singular or vicinal multilayer surface is considered for a GaAs and a Ag multilayer component layer. Previous results are extended to include line tension balances at the free surface. In all cases at 293K and for most cases at 1000K, nucleation does not occur for coherency strains below the theoretical failure strain of a perfect crystal. Nucleation occurs for the partial dislocation-vicinal surface case but only at very large misfit strains. The implication is that strained multilayer structures are stable against dislocation injection, even when the layer thickness exceeds the thermodynamic critical thickness for misfit dislocation stability, in the absence of defects or large external stresses. Hence, if a dislocation-free structure can be grown it should have enhanced stability relative to the thermodynamical stability limit. The results are lower bound calculations for critical thicknesses in the sense that we do not consider the resistance to dislocation motion once dislocations are nucleated. Lattice resistance, extrinsic defects, and frictional damping terms [15] would lead to shifts of critical thicknesses to larger values than those considered here.

An exact energy calculation is also presented for the spreading of threading dislocations in multilayer structures. The results show that in the absence of jog formation, there is no barrier to spreading at or above the thermodynamic Matthews-Blakeslee critical thickness h^*. Moreover, reversible spreading to an equilibrium standoff distance occurs for thicknesses smaller than h^*. This process can lead to dissipative reversed dislocation motion under alternating driving forces. With the requirement for jog formation in the initial stages there is a barrier to spreading for thicknesses in the vicinity of h^*.

ACKNOWLEDGEMENT

.

REFERENCES

1.   L. Esaki and R. Tsu, IBM J. Res. Dev. 14, 686 (1970).

2.   K. Ploog and G.H. Dohler, Adv. Phys. 32, 285 (1983).

3.   J.W. Matthews and A.E. Blakeslee, J. Cryst. Growth 27, 118 (1974).

4.   J.W. Matthews and A.E. Blakeslee, J. Cryst. Growth 29, 273 (1975).

5.   J.W. Matthews and A.E. Blakeslee, J. Cryst. Growth 32, 265 (1976).

6.   F.C. Frank, in "Symposium on Plastic Deformation of Crystaline Solids", (Carnegie Inst. of Tech., Pittsburgh, 1950) p. 89.

7.   J.P. Hirth, in "The Relation Between the Structure and Mechanical Properties of Alloys", (H.M. Stationere Office, London 1963), p. 217.

8.   J.W. Matthews, in "Dislocations in Solids", F. R.N. Nabarro, ed., Vol. 2, (North Holland, Amsterdam, 1979) p. 461.

9.   P.M.J. Marée J.C. Barbour, J.F. VanderVeen, K.L. Kavanagh, C.W.T. Bulle-Lieuwma and M.P.A. Viegers, J. Appl. Phys. 62, 4113 (1987).

10.   J.P. Hirth and A.G. Evans, J. Appl Phys. 60, 2372 (1986).

11.   J. Feder, K.C. Russell, J. Lothe and G.M. Pound, Adv. Phys. 15, 111 (1966).

12.   J.P. Hirth, and J. Lothe, Theory of Dislocations, 2nd ed. (Wiley, New York, 1982), p. 231.

13.   J.P. Hirth and J. Lothe, Theory of Dislocations, 2nd ed., (Wiley, New York, 1982) p. 243.

14.   A. Granato and K. Lücke, J. Appl. Phys. 27, 583 (1956).

15.   B.W. Dodson and J.Y. Tsao, Appl. Phys. Letts. 51, 1325 (1987).

Table 1. Values of $\frac{F}{f}$ and c_0^* for several cases of dislocation nucleation in a GaAs layer

	Perfect Dislocation				Partial Dislocation			
	293 K		1000 K		293 K		1000 K	
	$\frac{F}{f}$	c_0^*	$\frac{F}{f}$	c_0^*	$\frac{F}{f}$	c_0^*	$\frac{F}{f}$	c_0^*
Internal loop	0.97	0.31	1.31	0.26	1.51	0.43	2.64	0.30
Singular Surface	0.73	0.43	1.18	0.31	1.21	0.57	2.55	0.33
Vicinal Surface	2.85	0.12	3.60	0.11	7.47	0.11	11.0	0.087

Table 2. Values of $\frac{F}{f}$ and c_0^* for several cases of dislocation nucleation in a Ag layer

	Perfect Dislocation				Partial Dislocation			
	293 K		1000 K		293 K		1000 K	
	$\frac{F}{f}$	c_0^*	$\frac{F}{f}$	c_0^*	$\frac{F}{f}$	c_0^*	$\frac{F}{f}$	c_0^*
Internal loop	1.30	0.22	2.12	0.16	2.63	0.25	5.45	0.15
Singular Surface	1.12	0.27	2.20	0.17	2.40	0.29	5.70	0.15
Vicinal Surface	4.25	0.078	6.40	0.161	14.7	0.0565	25.2	0.041

FIGURE CAPTIONS

Fig.1   A pair of layers from a diamond cubic multilayer structure with a (001) interface.

Fig.2   Nucleation of a loop of a perfect $\frac{1}{2}$ [011](11$\bar{1}$) dislocation within a strained multilayer.

Fig.3   Nucleation of a loop of combined $\frac{1}{2}$ [011](11$\bar{1}$) dislocation at a singular surface in a strained multilayer.

Fig.4   Nucleation of a loop of combined $\frac{1}{2}$ [011](11$\bar{1}$) dislocation at a vicinal surface in a strained multilayer.

Fig.5   View parallel to the (001) interface of the spreading of a $\frac{1}{2}$ [011] dislocation on (11$\bar{1}$): (a) view perpendicular to $\underline{b}$ and ξ, (b) view parallel to the glide plane and (c) view as in (b) but with a dislocation with a jog E.

Fig.6   Critical configuration of dislocation loop when the radius exceeds $h'/2$.

Fig.7   Total energy (W) versus dislocation spreading length (a) for $\epsilon_o = 0.01$ and $h = 10$ nm.

Fig.8   Total energy (W) versus dislocation spreading length (a) for $\epsilon_o = 0.01$ and $h = 30$ nm

Fig.9   Total energy (W) versus dislocation spreading length (a) for $\epsilon_o = 0.01$ and $h = 70$ nm.

Fig.10   Equilibrium dislocation spreading length (a*) versus normalized layer thickness (h/h*) for $\epsilon_o = 0.01$.
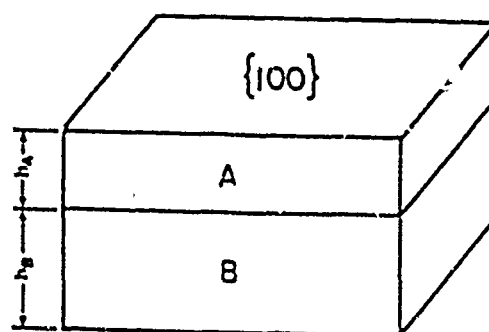
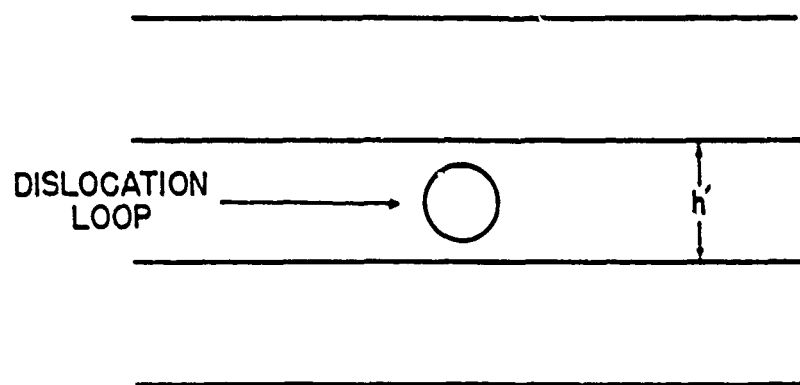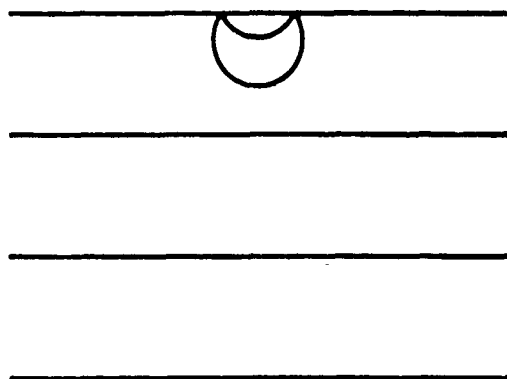Fig.1   A pair of layers from a diamond cubic multilayer structure with a (001) interface.

Fig.2   Nucleation of a loop of a perfect $\frac{1}{2}$ [011](11$\bar{1}$) dislocation within a strained multilayer.

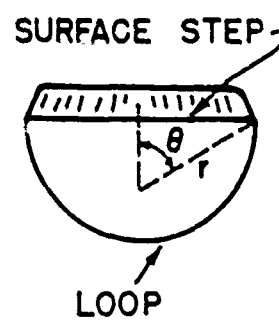a)   b)

Fig.3   Nucleation of a loop of combined $\frac{1}{2}$ [011](11$\bar{1}$) dislocation at a singular surface in a strained multilayer.

Fig.1   Nucleation of a loop of combined $\frac{1}{2}[011](11\bar{1})$ dislocation at a vicinal surface in a strained multilayer.
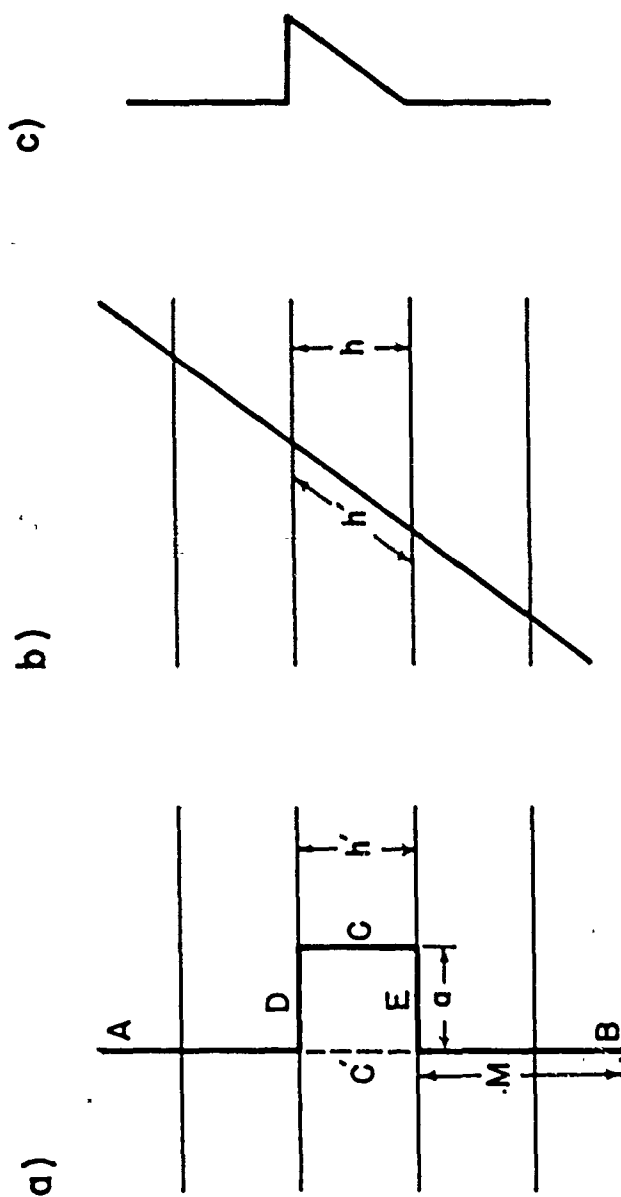
Fig.5 View parallel to the (001) interface of the spreading of a $\frac{1}{2}[0\bar{1}1]$ dislocation on $(11\bar{1})$: (a) view perpendicular to b and ξ, (b) view parallel to the glide plane and (c) view as in (b) but with a dislocation with a jog P.
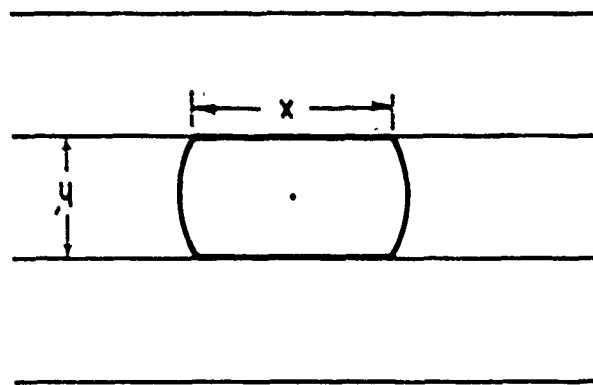
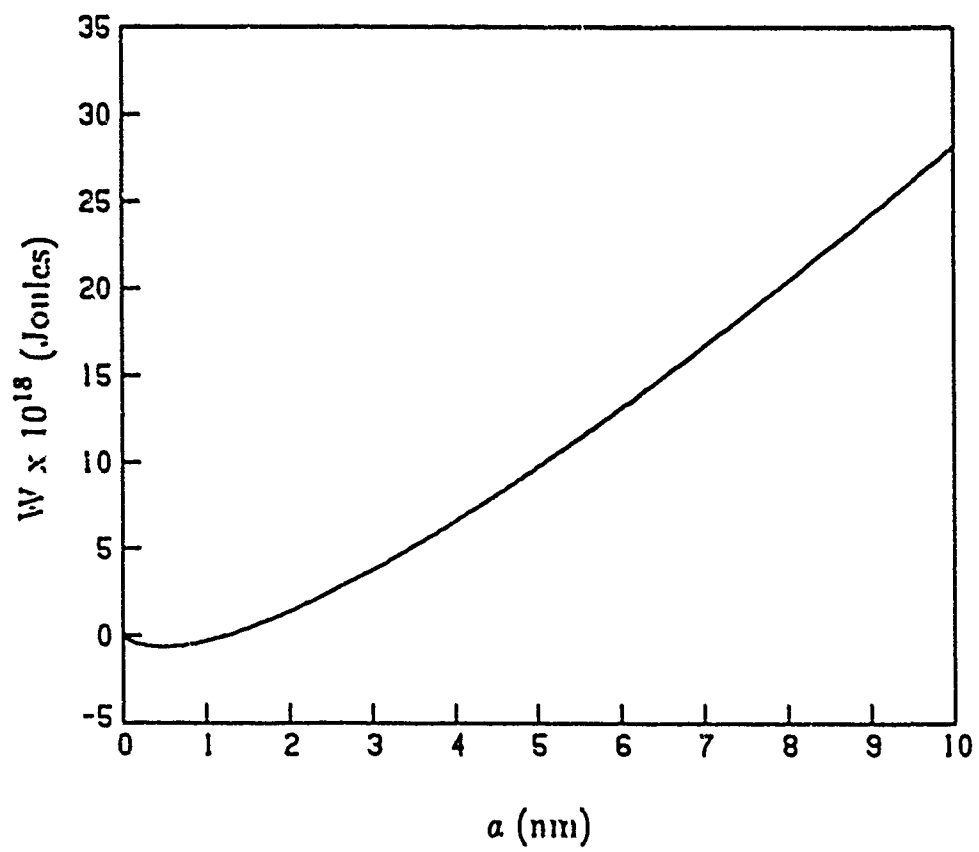Fig.6   Critical configuration of dislocation loop when the radius exceeds $h'/2$.

Fig.7   Total energy (W) versus dislocation spreading length (a) for $\epsilon_0 = 0.01$ and h = 10 nm.
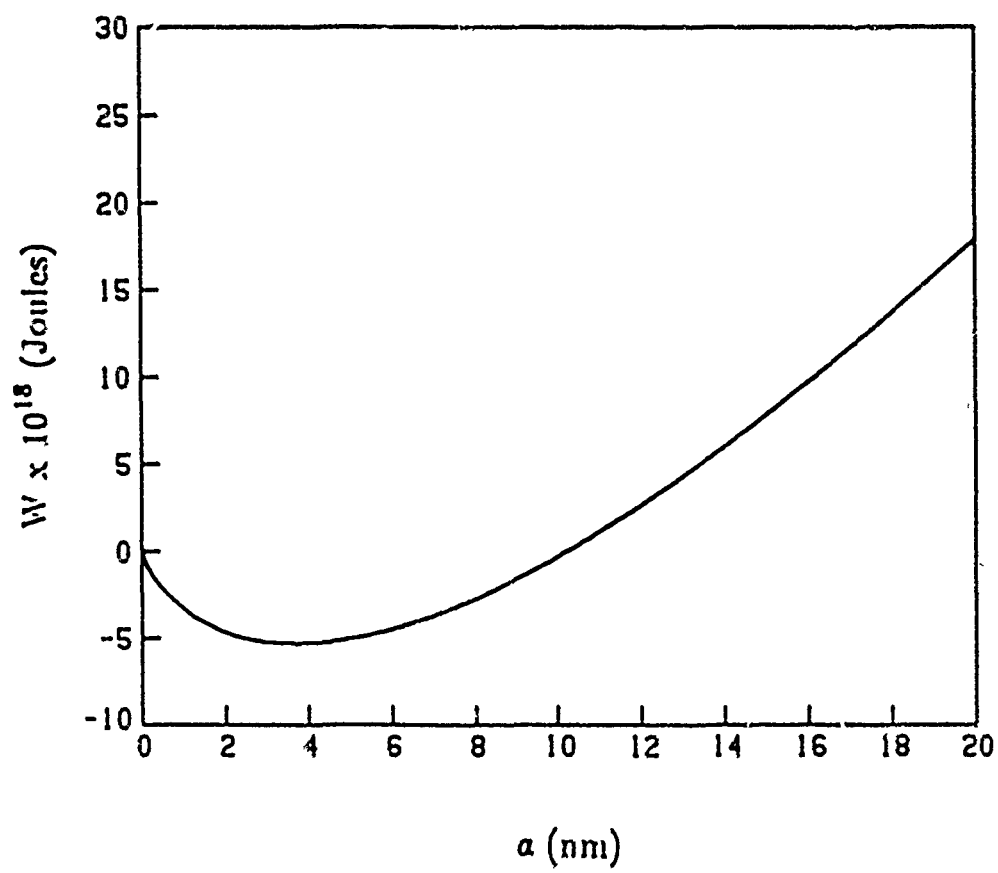
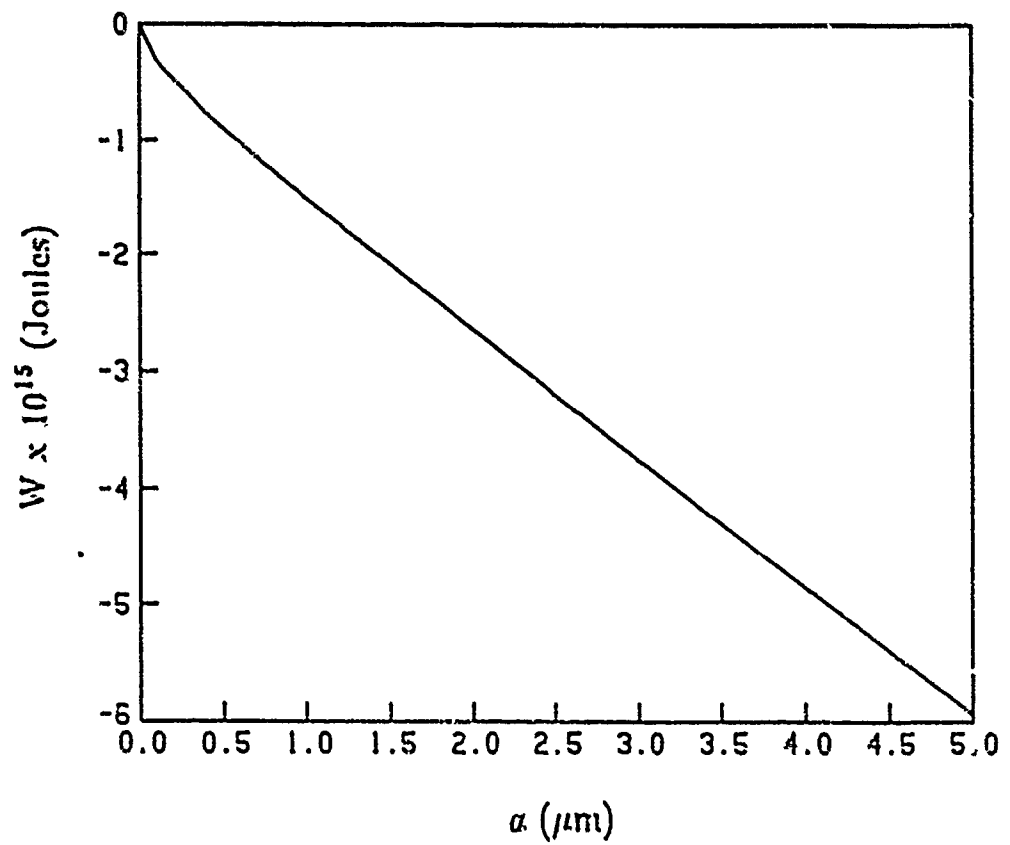Fig.8 Total energy (W) versus dislocation spreading length (a) for $c_\bullet = 0.01$ and $h = 30$ nm

Fig.9   Total energy (W) versus dislocation spreading length (a) for $c_o = 0.01$ and $h = 70$ nm.
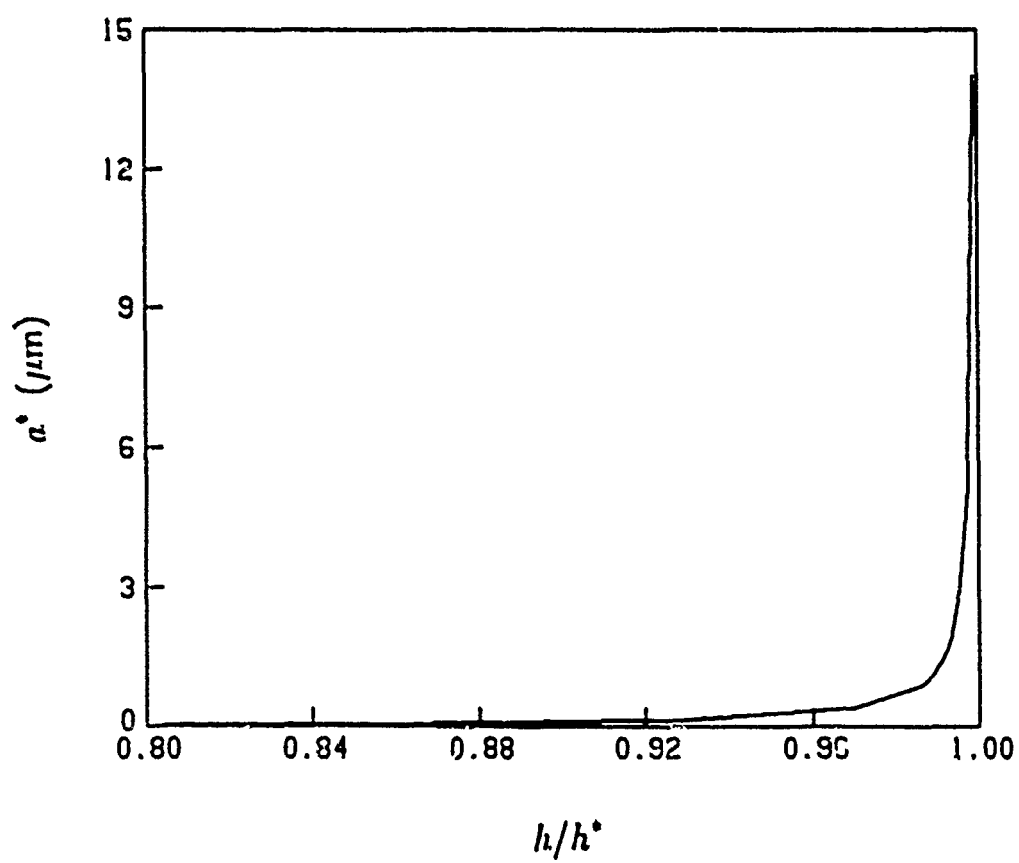
Fig.10 Equilibrium dislocation spreading length (a*) versus normalized layer thickness (h/h*) for c_o = 0.01.

MATERIALS

ON THE ATOMIC STRUCTURE
OF THE Nb/Al$_2$O$_3$ INTERFACE AND
THE GROWTH OF Al$_2$O$_3$ PARTICLES

M. Kuwabara and J. C. H. Spence

Department of Physics
Arizona State University
Tempe, Arizona 85287-1504

and

M. Rühle

Materials Department
College of Engineering
University of California, Santa Barbara
Santa Barbara, California 93106

* Now at Max-Planck-Institut für Metallforschung, Institut für
Werkstoffwissenschaft, Seestraße 92, D-7000 Stuttgart, FRG

On the atomic structure of the Nb/Al₂O₃ interface and the growth of Al₂O₃ particles

M. Kuwabara and J. C. H. Spence
Department of Physics, Arizona State University, Tempe, Arizona 85287-1504

M. Rühle
Materials Department, University of California–Santa Barbara, Santa Barbara, California 93106

The growth mechanism for small precipitates of Al₂O₃ formed by internal oxidation in the Nb–Al₂O₃ interface is studied in detail. The observations show that the Nb (001)//Al₂O₃ (00.1) interface is almost atomically flat and that there are no interface compounds. We suggest that the final layer on the Al₂O₃ side of this interface consists of oxygen atoms. The effects of image forces on misfit dislocations are found to result in a standoff distance between dislocation cores and the interface, in good agreement with the recent theory. The implications of this for the strength of metal-ceramic bonding are discussed.

I. INTRODUCTION

The study of the atomic structure at metal-ceramic interfaces is important for both fundamental and applied reasons. The niobium-sapphire system studied here has applications including Josephson junctions and structural ceramics, and has been widely studied as a model system.[1] Structural ceramics must be bonded to metals. The applicability of the ceramic depends mainly on the fracture resistance of the bond. An understanding of the adhesion between the dissimilar materials requires that structural details are available down to the atomic level. Recently it has become possible to prepare extremely well-characterized interfaces, and to study their structure at atomic resolution using high-resolution electron microscopy (HREM). This technique is now capable, in favorable cases, of determining the atomic structure of an interface of known composition.[4] It is therefore able to answer such questions as the location of misfit dislocation cores, the presence or absence of reaction layers, and the degree of surface roughness at interfaces (on an atomic scale). It is also possible in some cases to determine precisely the atomic structure at the interface plane (see Ref. 5 for a review).

In this paper we report a study of the growth mechanism of Al₂O₃ particles formed by internal oxidation in Nb. We also discuss the atomic structure of the Al₂O₃/Nb interface and find that, in agreement with recent elasticity calculations,[6] the misfit dislocations do not occur exactly at the interface. Rather, due to image forces, their cores are set back a small "standoff" distance into the Nb. The implications of this finding for the strength of metal-ceramic interfaces generally are discussed.

II. EXPERIMENTAL

The Nb–Al alloy was internally oxidized by holding niobium containing 3 at. % Al at 1450 °C for 40 min at a pressure of 5×10^{-3} mbar oxygen. This produces the Al₂O₃ precipitates, as described elsewhere.[2] The final specimens for the electron microscope were prepared by the argon ion bombardment of cut and dimpled samples. They suffer little radiation damage. However, because of the slightly different thinning rate of Nb and Al₂O₃, the thickness on the Nb side is always a little greater than that on the Al₂O₃ side [see Fig. 1(a)].

The electron microscope observation was performed in a thin region near the edge of the sample, using a JEOL-4000EX (operated at 400 kV). The wedge-shaped regions show the full variation of the image with thickness. The lattice images were obtained at the optimum focus condition, near Scherzer focus, under axial illumination.

III. RESULTS AND DISCUSSION

A. Structure of the interface

Several different types of Al₂O₃ precipitates were observed in the samples. Figure 1 shows an electron micrograph and the corresponding diffraction pattern of a typical Al₂O₃ particle which was frequently observed. In most cases the particles have a plate-like shape, as seen in Fig. 1(a). Figure 1(b) is a selected area diffraction pattern along the [100] Nb zone axis, including an Al₂O₃ particle. The strong diffracted spots from the Al₂O₃ particle are consistent with the α–Al₂O₃ structure along the [01.0] zone axis. The weak extra spots along the c* axis are not, however. Following Mäder,[7] this phase is denoted as α'–Al₂O₃. We consider it to be a four-layer stacking sequence variant of α–Al₂O₃. The interface shown in Fig. 1(a) is Nb (011)//Al₂O₃ (00.1), with Nb [100]//Al₂O₃ [12.0]. This orientation relation has lower energy than any other.[3] Because of the different thinning rate between Nb and Al₂O₃, the Nb side is always slightly thicker than the Al₂O₃ side. The thickness of Nb can be estimated from the thickness
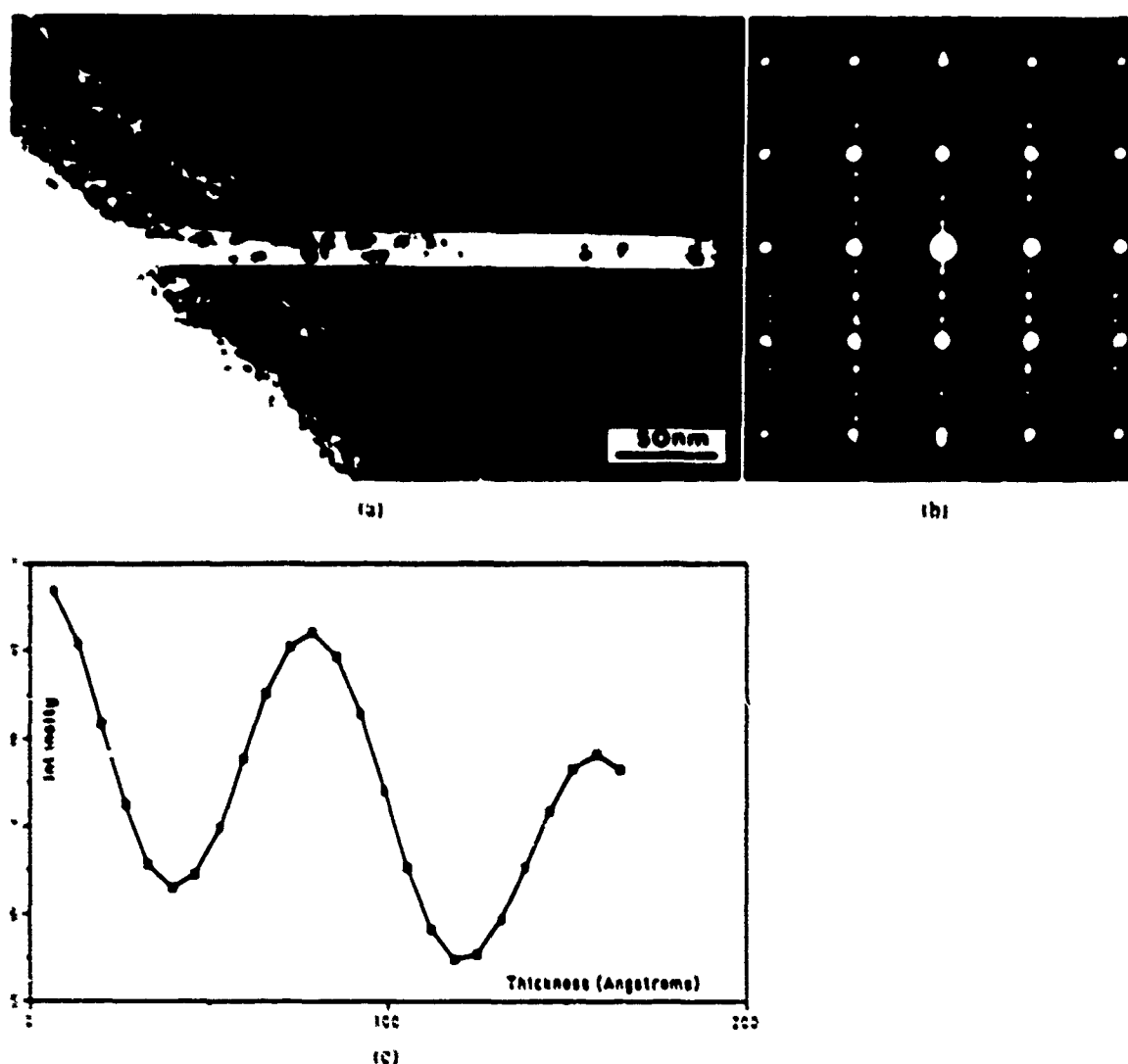
(a)

(b)

(c)

FIG. 1. (a) Electron microscope image of a Al_2O_3 precipitates in Nb crystal. (b) Selected area diffraction pattern of (a). The incident beam direction is the [100] Nb zone axis. The orientation relation at the interface is Nb ·[011], Al_2O_3 [00.1]. (c) Many-beam multislice calculation for the [000] beam in Nb, with electron beam parallel to [100] at 400 kV. Debye-Waller factor 0.3 and 100 beams have been included in the calculation. Doyle-Turner scattering factors.

fringes. The thickness in the first and second dark region is, respectively, 4 and 12 nm. This was determined by a comparison of the bright-field Nb images with many-beam Pendellosung calculations, as shown in Fig. 1(c).

Figure 2 shows a high resolution image of Fig. 1(a) near the crystal edge. The Nb thickness changes gradually from the edge to about 10 nm. The strong long-range contrast in the Al_2O_3 particle is due to a slight misalignment of crystal orientation. It is clear that this interface is atomically flat, and there are no reaction layers at the interface. This is a general result, found for all the interfaces studied. At the interface the lattice spacings of Nb and Al_2O_3 in this orientation are slightly different [$d(0\bar{1}1)$]

Nb = 0.2326 nm, $d(2\bar{1}.0)Al_2O_3$ = 0.2394 nm, and this value was measured from the diffraction pattern]. This difference is accommodated by misfit dislocations. In Fig. 2, the misfit dislocations can be clearly observed on both sides of the Al_2O_3 particles, as indicated by the arrows. They are arranged with approximately equal spacing. It is generally considered that the distance L between dislocations is determined by the difference in lattice spacing of the two crystals. To obtain an accurate value of L, the distance between dislocation cores was measured from the HREM images, and the average value found was L = 8.5 nm ± 0.2 nm. According to simple theory $L = d_1 d_2/(d_1 - d_2)$, (d_2 = 0.2326 nm, d_1 = 0.2394 nm). L is
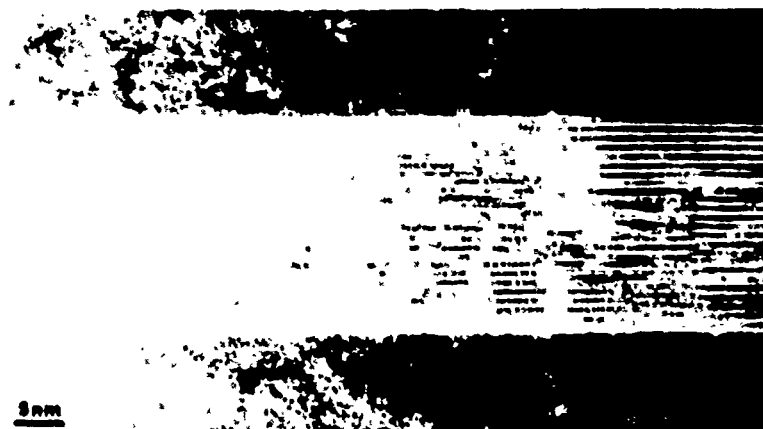
FIG. 2 HREM image of thin region imaged by Russian misfit dislocations arrowed with a set of high resolution

8.19 nm. The experimental value is thus somewhat larger than the calculated one. This may be due to relaxation in our very thin samples.

To understand the bonding mechanism at the Nb/Al$_2$O$_3$ interface it is important to determine which atom plane forms the last layer at the interface. This cannot be answered definitively without detailed image simulations for the interface. Despite simulations for several likely models for the interface, we have been unable to obtain a satisfactory match, due to thickness variations along and across the interface, and surface roughness. Al and O are bound by covalent forces so that the Al"O" Nb' arrangement seems to be reasonable on chemical grounds. In addition, for chemical reasons it can be expected that the outermost layer of Al$_2$O$_3$ will contain an excess of O", since Al$_2$O$_3$ is much more stable than Nb$_2$O$_5$, and any unsaturated Al bonds will be saturated with oxygen, which is present in excess during the internal oxidation process. Therefore we assume that the top layer at this interface is an oxygen atom plane.

A two-dimensional model for this interface is shown in Fig. 3. The top layer is an oxygen atom plane, and along the (10.0) direction the misfit dislocations observed in HREM images allow a periodic interface. However, at the interface the Nb and O atom planes have difference symmetry. As shown in the model, the Nb period L1 and Al period L2 along the (12.0) direction match well, since 2° (1.6436 nm) = 5°L1 (1.645 nm). This relationship keeps the Nb (011) and Al$_2$O$_3$ (00.1) planes parallel, even though they have different symmetry.

B. Standoff effect at the interface

Figure 4 shows the misfit dislocation core. These are dislocation loops, generated as the particles grow by, for example, mechanism (3) of Ref. 9 (see also Ref. 10). In this mechanism, a dislocation loop is generated around the end of the particle as required to accommodate mismatch as the particle grows. The requirement for this process is

that the self-energy of the loop plus the interaction energy of the loop with the precipitate be negative. This in turn provides a condition on the critical particle size for stable growth of the incoherent particle. Since the particle growth is competitive, we expect those incoherent particles which, by an accident of birth, have access to misfit dislocations to grow at the expense of smaller coherent particles. Very accurate thickness determination and detailed image simulations could be used to determine the details of the atom positions at the dislocation core, if it were assumed that these sessile dislocations are straight and unaffected by surface relaxation. For our purposes, however, the Scherzer focus images from the thinnest regions at 0.18 nm point resolution are sufficient to localize the dislocation cores to within about 0.2 nm. It is clear that the misfit dislocation consists of one extra half plane, and that its strain field extends about 4–5 layers along (011) from the core. Most of the core does not lie in the interface, but is located a few layers away from the interface. This averaged "standoff" distance, measured from this interface, is 0.54 nm.

The "standoff" distance d_i is a result of the balance between the image force (normally neglected) and adhesion forces due to bonding. The image force is analogous to the corresponding effect in electrostatics, and results from the requirement that stresses and displacements must be continuous across the boundary between material with different shear moduli μ_1 (for Nb) and μ_2 (for Al$_2$O$_3$). A similar effect is well known in the study of dislocations in metals approaching oxidized surfaces, where the oxide is found to repel dislocations. Since $\mu_2 > \mu_1$, the misfit dislocations are also repelled in our system from the interface into the softer material. A detailed analysis of the standoff distance for the Nb/Al$_2$O$_3$ interface has been published,[*] in which it is predicted that $d_i = 3.6$ interplanar spacings (using $\mu_1 = 28.7$ GPa and $\mu_2 = 179$ GPa), in good agreement with our observations.

To understand the effect of the standoff distance on the strength of the interface we consider a crack propagat-
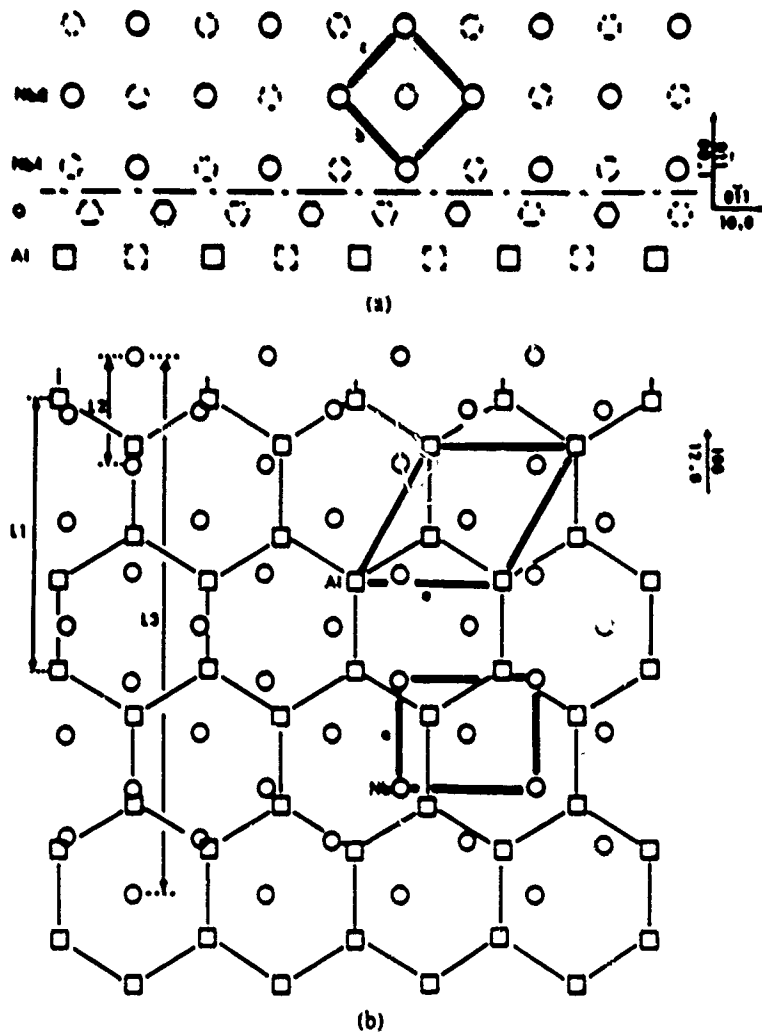
(a)

(b)

FIG. 3 Two-dimensional interface model of Nb (011)/Al₂O₃ (00.1) L1 and L2 represent, respectively, the Nb and Al period

ing along the Nb (110) interface. The effect of variations in *d*, may produce a shielding or antishielding effect on the crack tip, depending on the tip position. For the dislocation, we note that slip is commonly observed on the (110) interface plane [also (112) and (123) in bcc metals]. The force on the dislocation $F = b\sigma_{xy} = Kbf_{xy}(\theta)/(2\pi r)^{1/2}$ may be calculated for a pure mode I crack with stress-intensity factor K. (See Ref. 7 for plots of $f_{xy}(\theta)$ showing the θ regions in which a dislocation is attracted or repulsed from a crack tip in a homogenous medium.) A complete stress analysis of the problem of a dislocation near an interface between different media subject to forces from a crack has recently been completed.[8] However, our problem is complicated by the three-dimensional nature of the dislocation loops, and their behavior on the other particle faces. For an infinite interface, extensive numerical evaluation of the expressions given in Ref. 8 would appear to offer the only possibility for determining the effect of variations in d_t on crack propagation.

C. Al₂O₃ particle growth

The misfit dislocations on each side of the particles are almost symmetrically located across the particle, as seen in Fig. 2. This suggests that the misfit dislocations are loops around the particles, arising from the initial stage of particle growth. The mechanism for this loss of coherence in growing particles has been discussed extensively in the literature.[9,10]

As mentioned previously, most of the particles have a plate-like shape, and there are no steps on the Al₂O₃ basal plane. Figure 5 shows a different Al₂O₃ particle with the hexagonal symmetry of α-Al₂O₃ (corundum), the interfaces (A, C) between Nb (011) and Al₂O₃ (00.1) are also atomically flat. Moiré fringes caused by Nb and Al₂O₃ can be seen at the B and D interface; furthermore, there is a large step at D. This evidence indicates that the Al₂O₃ precipitates, which have α-like structure, tend to grow in directions normal to ư C axis of the Al₂O₃.

FIG. 4 Structure image of a misfit dislocation core. Note "standoff" distance between core and interfaces. The core does not lie exactly in the plane of the interface.

Figure 6 shows another typical Al_2O_3 particle. From the diffraction pattern, this particle is found to be γ-Al_2O_3. At the A and C interface many misfit dislocations can be observed; on the other hand, there are complicated dislocations and steps in the Nb crystal near the B and D interface. The dislocations are arranged symmetrically against the [001] plane indicated by the white lines, and the B, D interfaces are more rough than the A, C interfaces. Larger steps, similar to the one in Fig. 5, can be seen at the interface B. During the particle growth, oxygen atoms must be supplied from the surface. It is likely that these oxygen atoms enter at surface steps on the slip planes of the Nb crystal. We suggest that the dislocations observed at the B, D interface were formed during the particle growth step, with the particle growth along the indicated arrows. Growth activity is seen only at these ends of the particles.

IV. CONCLUSIONS

The present HREM observations lead to the following conclusions. The internal oxidation of Nb-Al alloy does not produce any intermediate compounds at Nb, Al_2O_3 interfaces. The interface between Nb (011) and α-Al_2O_3 (00.1) is atomically flat. The observed distance between misfit dislocations at the interface is larger than that calculated by simple theory. The outermost Al_2O_3 layer at this interface is believed to consist of oxygen atoms.

The standoff effect which has been predicted from theoretical work has been confirmed, and may influence the strength of the interface. This confirmation improves
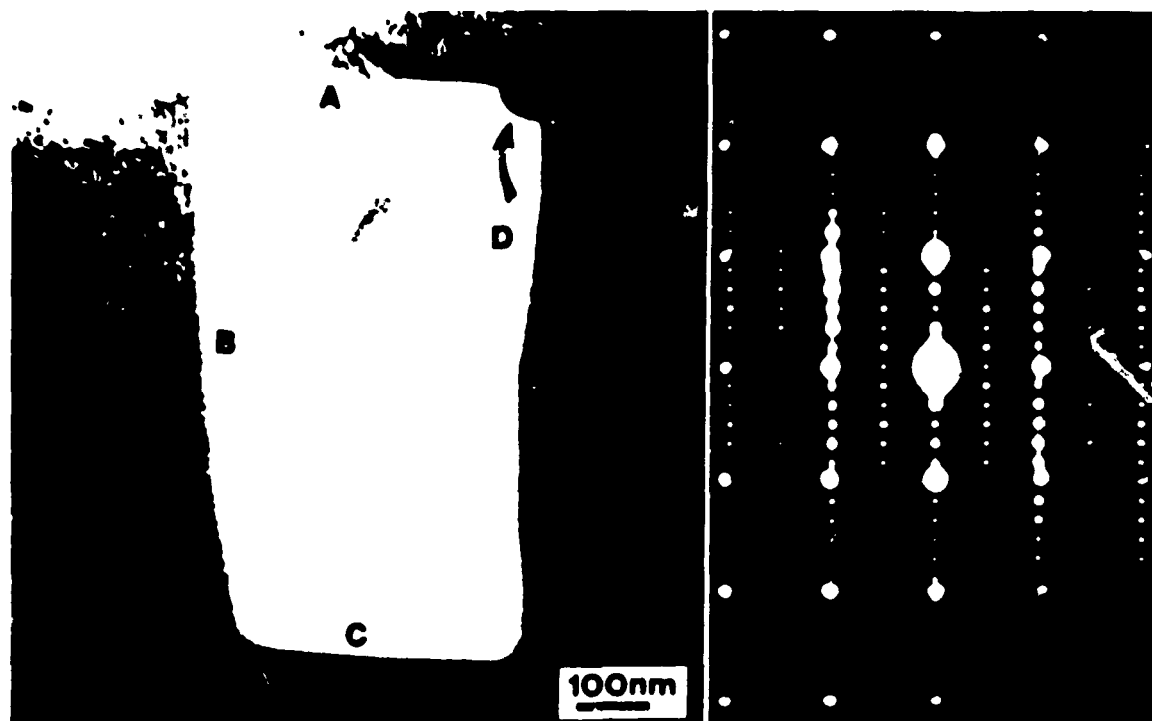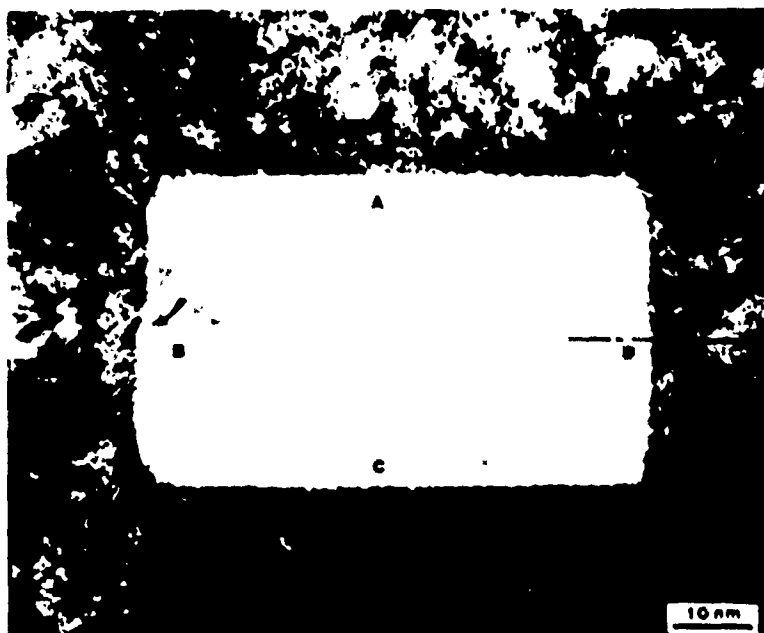

FIG. 5 Another α-like Al_2O_3 precipitate. Interface (A, C) Nb (011) and Al_2O_3 (00.1) are both atomically flat. Moiré fringes can be seen at B and D. The large step, which is responsible for the particle's growth, is on the interface D.

High Image of MgO precipitate. Mg In dislocation networks near interface X. Configuration of split dislocations of the may arise due to the oxygen supply from the surface.

our understanding of the metal/ceramics bonding mechanism and may assist in the design of stronger metal/ceramic interfaces.

It has also been suggested that oxygen atoms supplied from the crystal surface through dislocations on the slip planes of Nb. and Al₂O₃ particles grow by steps at the interface.

ACKNOWLEDGMENTS

REFERENCES

[1] M. Florjancis, W. Mader, M. Ruhle, and M. Turwitt, J. de Phys. (1986).

[2] K. Burger, W. Mader, and J. Ruhle, Ultramicroscopy 22 1 (1987)

[3] W. Mader, in Mat Res Soc Symp Proc (1986) (in press)

[4] D. Cherns, J C H. Spence, G R. Anstis, and J L. Hutchison, Phil Mag. A46, 819 (1982)

[5] See volume 14, No. 1 and 2 of Ultramicroscopy, devoted to electron microscopy of interfaces

[6] S V Kamat, J P Hirth, and B Carnahan, Mat Res Soc Symp Proc (Boston, MA, 1988), p 55 (The Materials Research Society, New York)

[7] B R Lawn and T R Wilsham, Fracture of Brittle Solids (C U P), 55 (1975)

[8] Z Suo, Int J of Solids and Structures (1989) (in press) See also Z. Suo, Proc Roy Soc (1989) (submitted)

[9] L M. Brown and G.R. Woodhouse, Phil Mag. 21, 329 (1970)

[10] R. G. Baker, D. G. Brandon, and J. Nutting, Phil. Mag. 4, 1339 (1959).

MATERIALS
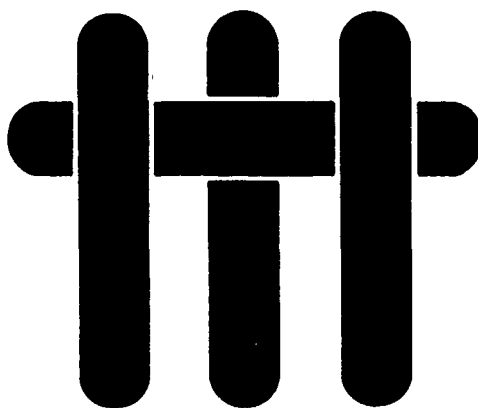
HREM STUDIES OF

Al$_2$O$_3$ PRECIPITATES IN Nb

M. Kuwabara, J. C. H. Spence
Department of Physics
Arizona State University
Tempe, Arizona 85287-1504

and

M. Rühle

Materials Department
College of Engineering
University of California, Santa Barbara
Santa Barbara, California 93106

* Now at Max-Planck-Institut für Metallforschung, Institut für
Werkstoffwissenschaft, Seestraße 92, D-7000 Stuttgart, FRG

HREM STUDIES OF Al₂O₃ PRECIPITATES IN Nb

M. Kuwabara*, J. C. H. Spence* and M. Rühle**

*Department of Physics, Arizona State University, Tempe, AZ 85787-1504
**Materials Department, University of California, Santa Barbara, CA 93106

Introduction   High resolution electron microscopy (HREM) studies are, in principle, capable of determining the atomic structure of defects, such as interfaces, in crystalline materials. A quantitative evaluation of the defect structure requires, however, that the examined specimen fulfills certain conditions, e.g. low-indexed zone axes of both materials adjacent to the interface must be parallel to each other and to the electron beam. Well defined geometrical characteristics exist between a metallic matrix and oxide precipitates formed by internal oxidation of a metallic alloy[1].

Results will be reported on structure and defects at Nb/Al₂O₃ interfaces formed by internal oxidation of a Nb-2.9 at % Al alloy (oxidation conditions: 30 min. at 1820 K, po₂ ~6.10⁻⁵mbar). In addition, information on growth mechanisms of the Al₂O₃ precipitate can be extracted from the HREM studies. An extended paper of the research will be published separately [2].

Structure of the Interface.   Most Al₂O₃ precipitates possess a plate-like shape with dimension of (200...600) x (200...600) x (10...40) nm³. The basis of the plate ("base interface") lies parallel to the (00.1) plane of Al₂O₃ and to a (110) plane of Nb with [001] Nb II [01.0] α'-Al₂O₃ [3]. No reaction layer could be identified at any interface. The base interface is atomically flat, no steps or facet could be identified. For a foil orientation parallel to (100) Nb the corresponding lattice planes (perpendicular to the interface) of Nb and Al₂O₃ possess different spacings (d [011]Nb = 0.2336nm, d[21.0]α'-Al₂O₃ = 0.2394nm.) The difference is accomodated by misfit dislocations (MFD) in Nb which can clearly be identified on both sides of the precipitate (Fig. 1).

Stand-off Effect at Interface   HREM studies show that the core of the MFD does not lie in the interface but is located a few layers away (Fig. 1, see also [3]). The average stand-off distance is in good agreement to theoretical predictions [4]. This "stand-off" distance is caused by a balance of interface coherency forces and image forces exerted on the MFD.

Growth of Al₂O₃ precipitates   The MFD on each side of the precipitate are almost symmetricaly located across the particle, (Fig. 2). This suggests that the MFDs are loops around the particles, arising from the initial stage of particle growth. The small precipitate shows atomically flat interfaces between (011) Nb and (00.1) Al₂O₃ at the interfaces A and C in Fig. 2.   Moiré fringes caused by an overlap of Nb and Al₂O₃ within the foil can be identified at the B and D interface. Large steps are found at B and D. The observations are indicative that the Al₂O₃ precipitates tend to grow in directions normal to the c- direction of Al₂O₃.

Conclusions.   Atomic structures of Nb/Al₂O₃ interfaces can be analysed by HREM. MFD are identified in Nb close to the interface. Information on the precipitate growth can be obtained by studying steps and facets at the interface [5].

References
[1] J. C. Meijering, in : Adv. Mat. Res. , Wiley 5 (1971) 1.
[2] M. Kuwabara, J. C. H. Spence and M. Rühle, to be published.

[3] W. Mader, Mat. Res. Soc. Symp. Proc. 82 (1987) 403.
[4] S. V. Kamat, J. P. Hirth and B. Carnahan. Proc. Symp. Mat. Res. Soc. 1988 in press.
[5] Work sponsored by DARPA, URI Program at UCSB, ONR Contract 00014-86-K-0753 and by ASU/NSF National Center for Electron Microscopy at Arizona State University.

Fig. 1

Fig. 2

Fig.1 HREM image of Nb/Al$_2$O$_3$ interface (internally oxidized Nb/Al alloy). Interface plane parallel to (0001) Al$_2$O$_3$ II (110) Nb and [01$\bar{1}$0] Al$_2$O$_3$ II [001] Nb. Misfit dislocations in Nb are arrowed. The extra half plane forming the misfit dislocation does not terminate at the interface.

Fig.2 HREM of small Al$_2$O$_3$ precipitate in Nb. Interfaces A and C are parallel to (011) Nb and (01.0) Al$_2$O$_3$ and are atomically flat. Steps and facets can be identified at interfaces B and D. Growth direction during oxidation.

MATERIALS
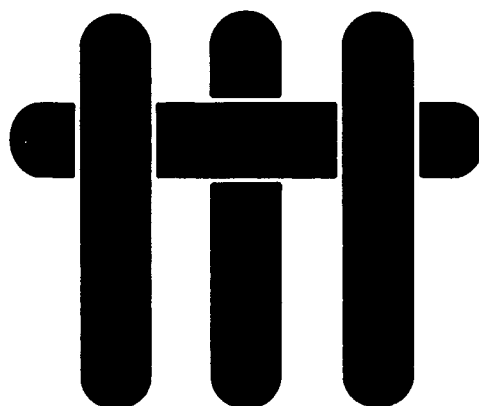
STRUCTURE AND CHEMISTRY OF

METAL/CERAMIC INTERFACES

*M. Rühle and A. G. Evans

Materials Department
College of Engineering
University of California, Santa Barbara
Santa Barbara, California 93106

* Now at Max-Planck-Institut für Metallforschung, Institut für
Werkstoffwissenschaft, Seestraße 92, D–7000 Stuttgart, FRG

187

Structure and Chemistry of Metal/Ceramic Interfaces*

M. RÜHLE and A. G. EVANS

Materials Department, College of Engineering, University of California, Santa Barbara, CA 93106 (U.S.A.)

(Received June 1, 1988)

Abstract

The present state of knowledge is reviewed concerning the structure and chemistry of metal/ ceramic interfaces. Experimental observations are described for several model systems and open problems concerning different aspects of structure and properties of heterophase boundaries are discussed.

1. Introduction

The use of ceramics as structural components, as well as in chemical technology and in electronic devices is steadily increasing because of improved mechanical integrity afforded by enhanced toughness and by process control. Ceramic components must typically be connected to other materials, mainly metals. The requirements that the bonded couple must fulfil are dictated by the functions of the ceramic: physical, chemical, electrical, mechanical. However, in all cases, adequate mechanical integrity is a technical prerequisite, as reflected in the fracture resistance of the interfaces. Metal/ ceramic bonded couples are presently being used in electron tubes, multilayer substrates and capacitors, metal matrix composites, automotive power sources, etc. [1-9].

Systematic studies of metal/ceramic interfaces started in the early 1960s. Such studies were directed towards the identification of general rules that govern bonding and interface behavior, both theoretically and experimentally, including the thermodynamics of interfacial reactions, crystallographic relationships and the atomistic structure at the interface. The intention of this article is to review the present state of knowledge concern-

ing the physics, chemistry and structure of interfacial regions between metals and ceramics.

2. The work of adhesion

The driving force for formation of a metal/ ceramic interface is the yield in energy when intimate contact is established between the metal and ceramic surfaces [10]. For a high rate of interaction, the surfaces have to be brought into excited states. Therefore, temperature and atmosphere are important variables, as well as the properties and structures of the surfaces.

The simplest description of the physical interaction between a metal and a ceramic is the *work of adhesion*, W_{ad}. Specifically, when clean, defect-free surfaces are brought into contact, energy is released in accordance with the Dupré equation:

$$W_{ad} = \gamma_c + \gamma_m - \gamma_{mc} \tag{1}$$

where γ_c and γ_m are the free energies of the relaxed surfaces of the ceramic and the metal, respectively. γ_{mc} represents the energy of the relaxed interface between the metal and the ceramic. The quantity W_{ad} is thus the reversible work released per unit area of interface formed by two free surfaces. Direct measurement of W_{ad} is not possible [11]. Consequently, in practice, W_{ad} is deduced by measuring the contact angle θ established by a *solid* metal in contact with a ceramic,

$$W_{ad} = \gamma_m (1 + \cos \theta) \tag{2}$$

Adequate measurement of θ and of γ_m constitutes a non-trivial experimental task. Often γ_m is anisotropic and, hence, the crystallography of the surface has to be determined. Furthermore, true equilibrium has to be established by allowing sufficient mass transport and the associated morphological evolution. Measurements on small particles are preferred, although contamination

---

*Paper presented at the symposium on Interfacial Phenomena in Composites: Processing, Characterization, and Mechanical Properties, Newport, RI, June 1-3, 1988.

188

during annealing is always a problem [12]. The most acceptable approach involves the deformation and heat treatment of alloys containing particles of ceramic formed by internal oxidation etc. [13, 14]. Plastic straining of the alloy causes particle decohesion. Subsequent annealing then allows mass transport to create an equilibrium void from the initial debond. The angle θ can then be measured on cross-sections through the particles [11]. Different authors [12, 14, 15] measured and calculated values of γ_m and W_{ad}.

Alloying additions strongly influence the thermodynamic quantities [10]. Furthermore, certain alloying additions segregate at the interface, by Gibbsian absorption. As an example, the segregation of chromium at various metal/Al_2O_3 interfaces results in a rearrangement of the interface into a more relaxed structure with a lower interfacial energy, resulting in a lower work of adhesion. Such segregant effects are a major issue in metal/ceramic bonded couples [12, 20, 21].

3. Bonding models

A rudimentary understanding of interfaces can be achieved by adopting phenomenological models. Such models are capable of correlating trends in bonding between different material couples and provide insight into some of the broad issues. However, the detailed understanding of trends in interface structure and properties with alloy composition, segregation, etc., requires more sophisticated atomistic models.

Elucidation of the essential issues, especially the prediction of trends in the work of adhesion (and, eventually, in fracture resistance) with such variables as alloy additions and segregation, requires that bonding be examined at all levels. The eventual objective would be the judicious coupling of information obtained from the most rigorous, but compute bound, quantum mechanical supercell approaches with the results of cluster calculations and of simple continuum, thermodynamic formalisms.

3.1. Continuum models

Interactions across the interface first occur without charge exchange. Such interactions develop between induced dipoles (London), between neutral atoms polarized by a dipole molecule (Debye) and between dipole moments (Keesom). (Another interaction without charge exchange involves anions in the ceramic and

image forces in the metal, and occurs when the dielectric constants of the metal and the ceramic are different [23-25].) Together, these interactions constitute the Van der Waals attractions (see for example ref. 22). The London term is generally the most pronounced. For a pair consisting of a metal atom and an oxygen ion, the interaction energy E_p has the form

$$E_p = -\frac{1}{2}\frac{\alpha_m\alpha_A}{R^6}\frac{I_mI_A}{I_m+I_A} \qquad (3)$$

where R is the distance between the centers of the interacting atoms/ions, α is the polarizability and I the ionization potential, with m referring to the metal and A the anion in the ceramic.

Charge exchange allows ion pairs to form and interact across the interfaces. For example, the interactions between ions of the metal and of oxygen (or other anions) in the ceramic is related to the free energy of metal oxide formation, ΔG^o [26, 27]. Furthermore, when the cations of the ceramic are soluble in the metal, dissolution from the interface allows ionic interaction between dissolved cations and the anions in the ceramic.

McDonald and Eberhart [28] examined interactions involving various metals in contact with the (0001) plane of sapphire. For this purpose, they assumed that the (0001) sapphire surface terminates with a layer of close-packed oxygen ions. The metal atoms (to be bonded to Al_2O_3) are then offered two sites: those above the aluminum ions located below the top layer of oxygen ions and those above empty sites. The first site results in attractive dipole forces, as described in eqn. (3), which are about constant for all metal/Al_2O_3 couples. The second metal site forms ionic oxygen-metal bonds having a free energy proportional to ΔG^o. By further assuming that all interactions of the dense-packed oxygen plane are occupied with metal atoms, the calculated trends in W_{ad} agree quite well with experimental data for the bonding of Al_2O_3 to simple metals [31]. The very strong bonding of platinum and palladium [19] to alumina is evidently at variance with the simple model. However, aluminum possesses a very high heat of solution in these metals. Consequently, as noted above, bonds could be formed between the oxygen ions and aluminum ions dissolved in the metal. Alternatively, a thin segregated aluminum layer could form between the metal and the Al_2O_3 to enhance the bonding [31].

The McDonald-Eberhart approach provides helpful generalizations. However, a more fundamental, atomistic understanding of the nature of the bonding is needed to adequately understand critically important alloying and segregation effects, as well as trends in the fracture resistance.

3.2 Atomistic models

An understanding of the fundamental physics of bonding between a metal and a ceramic requires that quantum mechanical models be developed. The simplest approach involves cluster calculations [30]. Such calculations have established that the primary interactions at metal/oxide interfaces involve the metal (d) and oxygen (p) orbitals, to create both bonding and antibonding orbitals. For copper and silver in contact with Al_2O_3, both states are about equally occupied, resulting in zero net bonding. However, for nickel and iron, fewer antibonding states are occupied and net bonding occurs. The calculations also reveal that a transfer of valence charge occurs, resulting in a contribution to the net ionic bonding which increases in strength as the metal becomes more noble. Consequently, metal-to-alumina bonding strengths are predicted to increase in the order: Ag-Cu-Ni-Fe. This order is generally consistent with the measured trends in sliding resistance as well as with the energies of adhesion. However, it is emphasized that the calculations approximate the interface by an $(AlO_x)^{y-}$ cluster and one metal atom. The selection of the charge to be assigned to this cluster is non-trivial and the choice influences the predicted magnitudes of the energies [32]. To further examine this issue, Anderson et al. [33] performed calculations for the Al_2O_3/Pt couple that included more atoms: 31 close-packed platinum atoms and the corresponding numbers of aluminum and oxygen ions. Then, by applying a quantum-chemical superposition technique, including an electron delocalization molecular orbital method, bonding energies were calculated for different atomic configurations of the Pt/Al_2O_3 interface. These calculations confirmed that the bond was strongest when oxidized platinum atoms opposed close-packed oxygen ion planes. However, further quantitative insights did not emerge. Indeed, the preceding models all have the deficiency that they do not fully account for the heterogeneous nature of the interface and cannot, therefore, be expected to accurately predict energies, segregant effects, etc.

Ab initio calculations seem to be essential for a full understanding of the bonding. Louie and coworkers [34, 35] have performed such calculations on metal-semiconductor interfaces. In these calculations, the metal was described by a jellium, so that insight emerged regarding the bonding mechanisms, but not on the atomistic structure. More recently, supercell calculations have been performed that include an interface area and adjacent regions large enough to incorporate the distorted (relaxed) volumes of both crystals. With this approach, the electron distribution around all atoms has been calculated and the atomic potentials evaluated. In a next step, interatomic forces may be calculated and strains determined. Such calculations have been performed rather successfully for the interface between Ge-GaAs [36] and Si-Ge [37]. The crystals adjacent to those interfaces are isomorphous and very nearly commensurate, such that the misfit between lattice planes is very small. However, misfits between metals and ceramics are typically rather large so that extremely large supercells are required. Therefore, only preliminary calculations have been conducted thus far [38]. Nevertheless, the calculations, performed for MgO/Ag, have allowed determination of the atomic potentials surrounding the different atoms, as well as a separation of bonding into different contributions (ionic, covalent and polarization).

With the advent of a new calculational scheme, Car and Parrinello [39] involving a combination of molecular dynamics (see e.g. Rahman [40]) and density functional theory (Kohn and Sham [41]), it should be possible to conduct computations of relaxed interfaces much more efficiently. The scheme should also allow equilibrium computations of metal/ceramic interfaces at finite temperatures. The conduct of such analysis on model interfaces should greatly facilitate the basic understanding of the bonding phenomenon and allow judicious usage of both cluster calculations and continuum thermodynamic formulations.

4. Structure of interfaces

In thermodynamic equilibrium, the atoms and/or ions close to an interface occupy positions that minimize the total energy of the system. However, the proximity to equilibrium depends on the conditions used to form the interface. For example, diffusion-bonded interfaces have atomistic arrangements influenced by the orientation of the

two surfaces *prior* to bonding. Additionally, when the bonding is performed at high temperature, chemical gradients often develop and influence the structure (see Section 5) and residual strains form upon cooling. The interface structure thus involves geometric as well as atomistic considerations, conditioned by relaxation mechanisms inherent in each bonding process. Consequently, interface structures are conveniently described by firstly defining generalized geometric parameters for the unrelaxed interfaces. Then, the relaxation mechanism pertinent to each bonding process may be considered. Finally, various geometric and atomistic bonding models, and associated experimental results, may be evaluated.

4.1. Geometrical parameters

An unrelaxed interface can be described by nine geometrical parameters. The required number was evaluated by a thought experiment, similar to that previously used for grain boundaries [42]. Six parameters describe the relative orientation and translation of the two crystals. The description of the interface orientation with respect to the crystal requires three additional parameters.

4.2. Process relaxations

The bonding method governs the actual geometrical parameters that describe the interface, by virtue of the imposed geometry and the allowable relaxations. During *diffusion bonding* [6], intimate surface contact at elevated temperatures, subject to a small pressure, generates the bonded interface. This technique, *pre-selects* the (macroscopic) rotation, two components of the translation and the interface orientation. However, some of these geometrical parameters are relaxed by *local* deviations. Specifically, bonding is usually performed at temperatures wherein at least one component may undergo plastic deformation. Consequently, local geometrical relaxation may be accommodated by small-angle grain boundary formation adjacent to the interface. Furthermore, mass transport may allow interface facets to develop that relax the constraint on interface planarity.

Interfaces may also be formed by *internal oxidation* [43, 44]. Such interfaces are not unique, but are related to the precipitate *morphology*, as governed by thermodynamic principles. In particular, since the total energy of the precipitate depends not only on the interfacial energy but

also on coherency strains. the interface structure and the shape of the precipitate depend usually on its size [45]. For coherent precipitates, orientation relationships are governed by constraints imposed by interfacial energies, the solubility and lattice parameters, and the most stable morphologies. These relationships should be an integral part of the analysis of the structure of interfaces formed by internal oxidation.

A third way of producing metal/ceramic interfaces is by *evaporation* of the metal onto a clean ceramic surface. The evaporated species is usually highly mobile [46] at high temperature. Consequently, if the metal wets the ceramic, a thin layer forms in an *equilibrium* interfacial configuration. For non-wetting configurations, islands are formed. Then, when the substrate is a single crystal and a one-to-one orientation relationship exists between the two components, the islands may grow together and form a single-crystal film [46]. Alternatively, when different equivalent orientation relationships exist, the islands develop with slightly different orientations. A range of different behaviors is thus expected for interfaces formed by evaporation.

4.3. Geometrical models

Models described by *geometrical* parameters may sometimes be insightful as a basis for the description of *unrelaxed* interfaces. However, it must be appreciated that the direct correlation between such models and interface properties (*e.g.* energy) is not possible [47]. Foremost among such models is a "lock-in" model developed by Gleiter and Fecht [48], deduced from experiments wherein small metallic single crystals were sintered onto a single-crystal ceramic substrate. Interfaces between several alkali halides or rock salt structure oxides and various noble metals were formed in this manner. It was observed that the spheres rotated into orientations which tend to minimize the energy of the sphere/plate boundary. Consequently, the orientation relationships determined by X-ray diffraction may be supposed to reflect a low-energy configuration of those interfaces. Gleiter and Fecht [48] accounted for the observed "stable" configurations by means of a model which describes low-energy interfaces in terms of densely packed atom rows of one crystal nesting in the grooves between similar rows of the other crystal. The requisite periodic matching between (small) multiples of lattice plane spacings was achieved by imposing small

displacements upon certain of the atom rows. The lattice strains from such imposed displacements have thus far been neglected. Nevertheless, in some cases, "lock-in" seemingly describes observations, even when large misfits exist. Conformity with the "lock-in" structure has been established for MgO/Au interface formed by evaporation[49] and for gold on KCl [35% misfit] [48]. In both cases, the predicted cube-on-cube orientation was observed. The "lock-in" concept thus appears to be a useful first-order description in some cases. However, more detailed investigations have revealed important discrepancies. Various observations indicate that small but definite angular off-sets occur between close-packed directions and/or planes. Mulder and Klomp [50] studied orientation relationships in diffusion-bonded sandwiches of polycrystalline copper and platinum foils between identically oriented (parallel) sapphire crystals. As expected, the metal grains developed a preferred orientation, with ⟨111⟩ of the metal parallel to ⟨0001⟩ of the sappi, e. However, the grains were rotated within a wide angular range about the close-packed directions. Discrepancies have also been found for the Nb/Al₂O₃ system [51, 52]. Notably, the close-packed planes are not aligned: instead, the (0001) plane of Al₂O₃ and (011) of Nb are tilted by about 3° around an axis parallel to [1120] Al₂O₃ and (101)ₙᵦ. This tilt occurs in specimens grown by epitaxy of niobium on (0001) Al₂O₃ [51] and in specimens formed by diffusion bonding [52], even when initially the specimens were adjusted to have parallel close-packed planes. It is also noted that even though MgO/Au has the cube-on-cube orientation predicted by "lock-in", misfit dislocations are observed, indicative of interface strains. Such dislocations are not consistent with the "lock-in" concept.

Relaxation effects are presumably responsible for discrepancies with the "lock-in" model. For example, as noted above, subgrain boundaries in diffusion-bonded Nb/Al₂O₃ provide freedom for some rotation of the niobium lattice. Also, the interface develops facets while retaining the 3° tilt of the two lattices. Clearly, the facets alleviate the misfit strain which would develop if close-packed planes were kept in contact everywhere across the interface.

4.4. Atomistic models

The *relaxed* structure of interfaces is obviously of greater significance than the unrelaxed state.

Adequate models of this structure do not yet exist, but await comprehensive experimental insight. Experimental studies can be performed either by diffraction [53, 54] or by high-resolution (direct imaging) techniques [55].* With the new generation of electron microscopes, direct imaging of *atomic columns* is possible. Specifically, by aligning along the axis of an interface which possesses only a tilt component, each column of atoms is imaged in one spot, giving a projection of the atomic arrangements in the interface. However, to derive reliable atomic positions, a series of observed images, taken at different focus settings, must be matched quantitatively with computed images, based on an assumed set of atomic coordinates. For matching purposes, image computation and position adjustment is repeated until the best possible fit is reached for the entire through-focus series. This technique is referred to as quantitative high-resolution electron microscopy (HREM) [56]. The interpretation is, of course, most difficult close to the interface where deviations from the perfect lattice are most extreme. Fully quantitative HREM studies have been performed on grain boundaries in germanium [55] and on semiconductor/metal interfaces such as Si/NiSi₂ [57].

Quantitative studies of metal/ceramic interfaces have been initiated for the Nb/Al₂O₃ and Au/MgO systems formed using each of the three processes described above. Interpretation of high-resolution images of regions close to a *diffusion-bonded* interface [58] (Fig. 1) requires that image simulations for perfect lattices first be performed. For this purpose, the projected atomic positions and potentials of niobium and Al₂O₃ in the selected orientation have been determined and used to construct the simulated image for a foil having a thickness (12 nm) which corresponds with that for the actual test specimen. It is evident that lens aberration ($C_s = 1.1$ mm, 200 kV) conceals some details of the perfect lattice structure. Nevertheless, it is still apparent (Fig. 1) that the perfect lattice is preserved in the Al₂O₃ up to the interface. Conversely, in the niobium strong deviations occur for distances up to four lattice planes from the interface. However, a periodicity along the interface can be recog-

---

*Diffraction studies require specimens with interfaces that extend homogeneously over a large area (about 1 mm²), while direct lattice imaging requires a thin specimen (thickness < 10 nm) plus constraints concerning homogeneity and orientation.

Fig. 1 Lattice image of the region close to an interface between niobium and Al₂O₃. (Orientation relationship between the ceramic and the metal: (0001) Al₂O₃ ∥ (110) Nb and [2̄110] Al₂O₃ ∥ [001] Nb. The foil orientation is parallel to [001] Nb. Insets in the high resolution electron microscopy images of niobium and Al₂O₃ represent simulated lattice images. Only four lattice planes within the niobium next to the interface are distorted. Facets can be observed. The periodicity of the interface is indicated.

Fig. 2. Lattice image of Nb/Al₂O₃ interface of an internally oxidized Nb/Al alloy. Interface plane parallel to (0001) Al₂O₃ ∥ (110) Nb and [01̄10] Al₂O₃ ∥ [001] Nb. Misfit dislocations in niobium can be observed (see inset). The extra half plane forming the misfit dislocation does not terminate at the interface (S).

nized within the distorted region. Faceting parallel to the interface also occurs and the interface is slightly inclined with respect to the (0001) plane of Al₂O₃. It is expected that imaging of the same interface with instruments possessing superior resolution would result in a better resolution of the distorted atomistic structure close to the interface. When the results of such studies become available, insightful atomistic models of the interface might be developed.

Interfaces created by *internal oxidation* have well defined geometrical characteristics. Precipitates of α-Al₂O₃ formed in Nb/Al alloys are small and penny-shaped (about 300 nm in diameter) and exhibit a fixed orientation relationship:

$$(0001)_{Al_2O_3} \| (110)_{Nb}$$

and

$$[01\bar{1}0]_{Al_2O_3} \| [001]_{Nb}$$

High-resolution transmission electron microscopy (TEM) investigation (Fig. 2) has revealed [59], however, that the spacings of the (110)Nb planes ($d = 0.23$ nm) and of the (0001) Al₂O₃ planes ($d_3 = 0.39$ nm) are sufficiently different that misfit dislocations would be predicted with spacing $D = d_1 d_3/(d_3 - d_1) = 8.8$ nm. Weak-beam images confirm such dislocations (Fig. 3) with $D = 9.1 \pm 0.5$ nm. To fully analyze the interface structure, as well as the atomistic structure of the
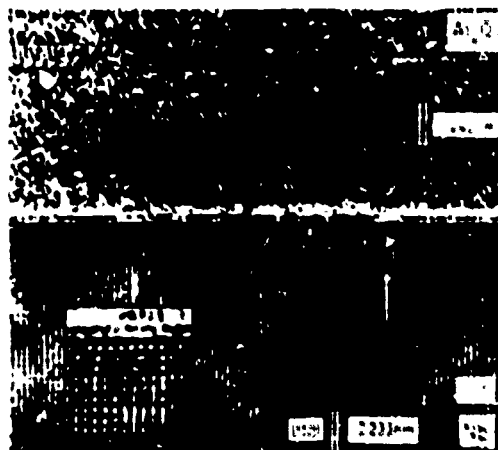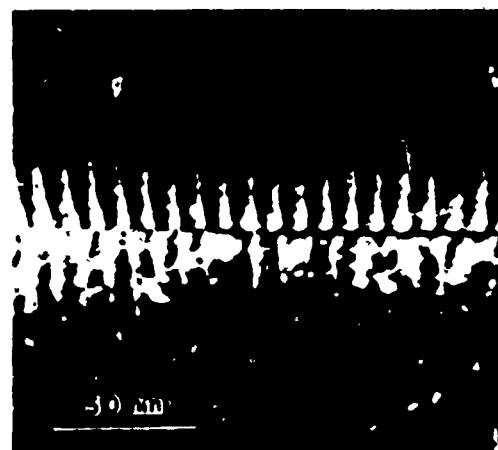

Fig. 3. Weak-beam dark-field image of inclined interfaces of Fig. 2 limiting the α-Al₂O₃ precipitate. Imaging with Nb (01̄0) reflection. Light lines represent the contrast from misfit dislocations.

misfit dislocations, superior resolution (less than 0.2 nm) would be needed.

Finally, it is noted that conventional TEM provides complementary information on defects at or close to the interface, such as dislocation spacings, ledge densities, etc. It is emphasized, however, that a more complete analysis of such defects, including Burgers vector determination, is generally not possible since the materials adjacent to the interface possess (usually) dif-

ferent lattice parameters and few "common" diffraction vectors exist [60].

5. Chemistry of interfaces

5.1. Theoretical background

In multicomponent two-phase systems, nonplanar interfaces or two-phase product regions can evolve from initially planar interfaces [61], even at constant temperature and pressure. Under the same conditions, the interface stays planar in binary systems. This difference originates with the thermodynamic degrees of freedom, f. For a binary system $f = 0$, whereas for ternary or higher-order systems, $f > 0$ (Gibbs' phase rule). Consequently, in the latter, interface compositions are in part controlled by the kinetics.

Not all (higher-order) interfaces necessarily develop an unstable morphology during a high-temperature treatment. Analysis of the phenomenon is needed to assess susceptibility. Similar problems exist for solidification and for the oxidation of alloys [62]. Mathematical treatments predict the time evolution of small interface perturbations. The perturbations may occur either due to initial roughnesses or upon small transport fluctuations caused by changes in temperature or by the presence of defects. If perturbations increase in amplitude with time, initially planar interfaces become morphologically unstable. The critical conditions for instability depend primarily on the mobility of the constituents and the thermodynamic properties of the system.

The formalism previously developed for ternary systems [61] can be adapted to metal/ceramic couples, with the three independent components being the two cations and the anion. A schematic ternary phase diagram and the expected concentration profiles are shown in Fig. 4 [63]. In general, the problem is complicated by having several phase fields present, such that intermediate phases form: usually intermetallics with noble metals and spinel (or other oxides) with less noble metals. The actual phases depend on the geometry of the tie lines, as well as on the diffusion paths in the ternary phase field, and cannot be predicted a priori.

The diffusion problem has thus far been examined [63] for the simple case wherein no product phases formed, the interfacial stresses were negligible, mass transport occurred by bulk diffusion and local equilibrium was imposed
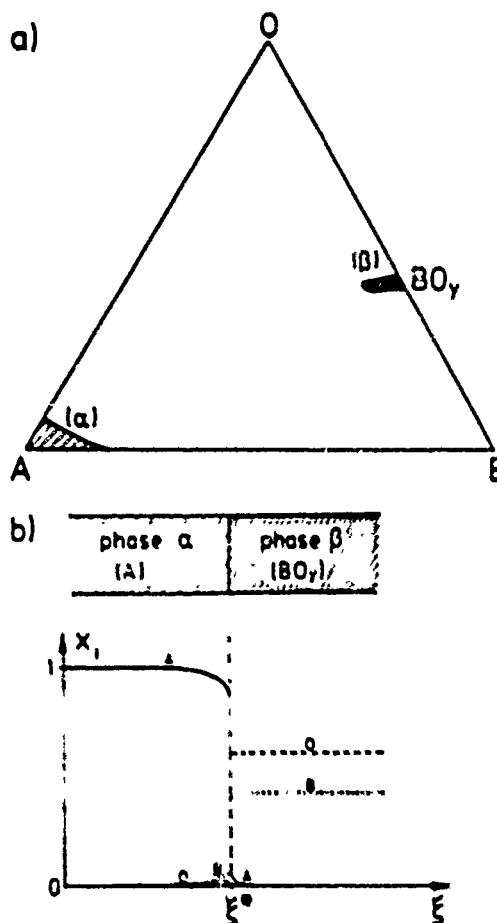

Fig. 4. (a) Schematic ternary phase diagram. Extended phases fields exist near metal A and oxide BO_x. (b) Equilibrium concentration profiles during diffusion bonding between metal phase (A) and oxide BO_x.

everywhere. Even then, a general analytical solution was not possible. However, for several metal/ceramic systems, some further simplifications are appropriate. The oxygen and the metal atoms diffuse on different sublattices allowing the interaction term in the diffusion coefficients to be neglected. (In the niobium, oxygen diffuses on interstitial sites while the aluminum diffuses by vacancies.) Negligible solubility of the metal in the oxide (grad $\mu_{MO} = 0$) allows point defect relaxation only in the ceramic. Therefore, for a stoichiometric ceramic, the remaining defect fluxes are small compared with the fluxes in the metal and can be ignored. Subject to the above simplifications, solutions have been obtained for Nb/Al_2O_3 [63, 64].

Rather high and probably unrealistic values of solute concentrations are calculated at the interface. In particular, the aluminum concentration profile in the niobium is steep and reveals an enrichment very near the interface, whereas the oxygen concentration profile is more extended, but with small absolute values. Such results do not agree with experimental observations, as elaborated below, suggesting that several assumptions are invalid. More sophisticated numerical models will thus require development as additional experimental results become available.

5.2. Interfaces without reaction layers

Detailed scanning electron microscopy (SEM) and TEM studies performed for Nb/Al₂O₃ [64–67] have shown that no reaction layer forms (see Fig. 1). Concentration profiles determined on cross-sections of rapidly cooled specimens revealed that, close to the interface, the concentration of aluminum is below the limit of detectability. However, with increasing distance from the interface, the concentration of aluminum, c_{Al} increases to a saturation value, $c_{Al}^* = 0.75$ wt.%, at a distance of about 2.5 μm (Fig. 5). The c_{Al} remains at that level up to a distance of $d^* \approx 16$ μm after bonding for 2 h and then decays exponentially. The magnitude of d^* depends on the bonding time. The corresponding oxygen content is below the limit of detectability. These measurements suggest that at the bonding temperature, c_{Al}^* at the interface possesses a value governed by the solubility limit. This limiting concentration would then extend into the niobium to a characteristic distance that depends on the

Fig. 5. Measured concentration profiles of aluminum in niobium as a function of the distance of bonding time and of the interface. Dih i, bonding conditions: 2073 K, 2 h, dynamic vacuum (orr), "fsst" cooling 215 K min⁻¹. (a) distances 150 i· listances 0–20 μm (higher spatial resolution).

bonding temperature and time. Upon cooling, the solubility of aluminum in niobium decreases, causing some of the dissolved aluminum (as well as oxygen) to diffuse back to the interface and condense as Al₂O₃.

"Slow" cooling after bonding resulted in completely different observations. Instead, small precipitates of θ-Al₂O₃ form in the niobium at distances between 8 μm and 14 μm from the interface [65, 66]. Furthermore, close to the zone wherein precipitation occurs, c_{Al} is very small. The precipitation presumably occurs during slow cooling, because the time requirements for precipitate nucleation are satisfied.

Bonding between platinum and Al₂O₃ subject to an inert atmosphere also occurs without chemical reaction. Specifically, no aluminum can be detected by Auger spectroscopy on the platinum side of an interfacial fracture surface [18, 19]. However, for bonds formed subject to a hydrogen atmosphere containing about 100 ppm H₂O, aluminum is detected in the platinum, indicative of Al₂O₃ being dissolved by platinum. More detailed studies concerning local chemical compositions are clearly required for a better understanding of the bonding processes involved.

5.3. Interfaces with reaction layer

For systems that form interphases, it is important to be able to predict those product phases created during diffusion bonding (given the possible phases present in the phase diagram). However, even if all the thermodynamic data are known, so that the different phase fields and the connecting tie lines can be calculated, the preferred product phase still cannot be unambiguously determined. The problem involves kinetic considerations. Specifically, the diffusion paths in phase space are controlled by different diffusion coefficients and, consequently, interface compositions depend also on the diffusivity ratios. Sometimes, small changes in the initial conditions can influence the reaction path dramatically, as exemplified by the Ni-Al-O systems [68]. Under high vacuum conditions (activity of oxygen $a_0 < 10^{-12}$) the diffusion path in the extended nickel phase field follows that side of the miscibility gap rich in aluminum and low in oxygen, (path I in Fig. 6), caused by the more rapid diffusion of oxygen than aluminum in nickel. This interface composition is directly connected by a tie line to the Al₂O₃ phase field, such that no product phase forms. However, whenever nickel contains suf-
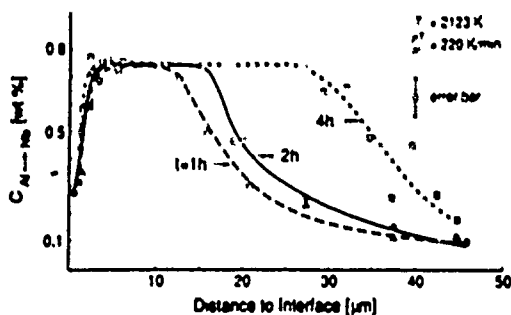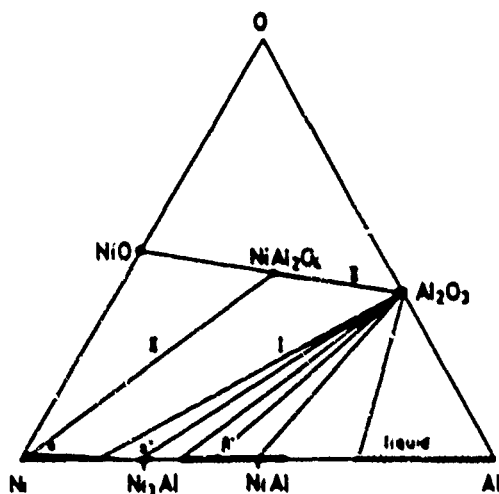
Fig. 6. Ni-Al-O phase diagram, schematically for T=1600 K. Two reaction paths are possible when nickel is bonded to Al_2O_3: (I) Low oxygen activity: no reaction product forms. (II) high oxygen activity: spinel forms.

ficient oxygen (about 500 ppm solubility), the $Ni(O)/Al_2O_3$ diffusion couple yields a spinel product layer. Under these conditions, spinel forms, because the new diffusion path in the nickel phase field requires that the tie line connects the metal and spinel field (path II in Fig. 6). The associated thermodynamic and atomistic consideration pertinent to spinel layer formation have been addressed [67, 68]. However, available observations do not unequivocally identify the operative mechanism. It is also noted that the interface between spinel and nickel seems to be unstable, morphological instabilities becoming more apparent with increasing spinel layer thickness.

Bonding of copper to Al_2O_3 seems to require a thin layer of oxygen on the surface of copper prior to bonding and $CuAl_2O_3$ or $CuAl_2O_4$ form [69]. The spinel thickness can be reduced by annealing under extremely low oxygen activities leading first to a "non-wetting" layer of Cu_2O and then to a direct Cu/Al_2O_3 bond. The mechanical stability of spinel-free Cu/Al_2O_3 specimens has not yet been investigated.

Bonding of titanium to Al_2O_3 results in the formation of the intermetallic phases TiAl or Ti_3Al, which probably also include oxygen. The thickness of the reactive layer increases with increasing bonding time and morphologically unstable interfaces develop. A detailed study is

again required for the identification of the different stable phases.

Similar studies have been performed for other ceramic partners such as simple cubic oxides (MgO, NiO, ...), sesquioxides (Cr_2O_3, Mn_2O_3, ...), Si_3N_4 and SiC. The situation is much more complicated for the latter materials, since impurities or sintering additives quite frequently diffuse to the interface and form a glassy film. The bonding is then governed by the interfaces between the glass film and both the ceramic and the metal.

6. Sintering of interfacial flaws

Scratches and other defects frequently exist on surfaces which have to be bonded. Long bonding times and high bonding temperatures are required when such flaws are present, because residual defects at the interface restrict the mechanical properties of the metal/ceramic interface. Investigation of the sintering of interface flaws is thus of great importance. The closure of interfacial flaws at homogeneous bonds involves several mechanisms [70]: surface diffusion, volume diffusion, diffusion along the interface, power-law creep and plastic yielding. Models for homogeneous interfaces based on these mechanisms qualitatively describe the experimental observations. However, a quantitative comparison has been inhibited by the scarcity of reliable experimental data. For a heterogeneous bond and especially for a metal/ceramic interface, additional processes have to be considered [64]; namely, the diffusion of metal atoms into the ceramic, dissolution of the ceramic at the interface and diffusion of the species of the ceramic into the metal, chemical reactions between the metal and the ceramic to form a product phase and recondensation of a dissolved ceramic at the interface. The inherent complexity demands a vital role of experiment in ascertaining the most important variables for typical metal/ceramic couples.

For evaluation of the governing transport mechanisms, artificial interfacial defects of different dimensionality and sizes may be introduced into the metallic surfaces [66, 67] prior to bonding: linear flaws by photolithography and other shapes by indentation. After bonding, cross-sections of interfacial fracture surfaces studied by SEM allow determination of the change in flaw size. The depth variations may be measured with a profilometer.

388

Inspection of results obtained for Nb/Al$_2$O$_3$ revealed that the surface of niobium hardly changes, whereas condensation products of Al$_2$O$_3$ could be detected on the sapphire. The amount of condensed Al$_2$O$_3$ depends on the cooling. From chemistry studies (Section 4), it has already been established that niobium dissolves Al$_2$O$_3$ and, upon cooling, Al$_2$O$_3$ recondenses at the interface. This mechanism also seems to be involved in the "filling" of the flaws. However, the experimental, as well as the theoretical, studies are not yet sufficient to derive a quantitative description of the process. Studies of this type on other systems are also needed to obtain a comprehensive view of the important issues in diffusion bonding.

7. Conclusions

Although considerable progress has been made in the understanding of structure and chemistry of metal/ceramic interfaces, still many open questions exist. Detailed and careful experimental   will reveal more insight into the mechai   bondings and these experimental studies r   allenge further theoretical studies.

Acknowledgment

References

1 E. A. Giess, K. N. Tu and D. R. Uhlmann (eds.), Electronic Packaging Materials Science, Mater. Res. Soc. Symp. Proc., 40 (1985).
2 K. A. Jackson, R. C. Pohanka, D. R. Uhlmann and D. R. Ulrich (eds.), Electronic Packaging Materials Science II, Mater. Res. Soc. Symp. Proc., 72 (1986).
3 J. M. Gibson and L. R. Dawson (eds.), Layered Structures, Epitaxy and Interfaces, Mater. Res. Soc. Symp. Proc., 37 (1985).
4 A. K. Dhingra and S. G. Fishman (eds.), Interfaces in Metal-Matrix Composites, The Metallurgical Society of AIME, Warrendale, PA, 1986.
5 M. Ruhle, R. W. Balluffi, H. Fischmeister and S. L. Sass (eds.), Proc. Int. Conf. on the Structure and Properties of Internal Interfaces, J. Phys. Paris, Colloq., C4-46 (1985).
6 Y. Ishida (ed.), Fundamentals of Diffusion Bonding, Elsevier, Amsterdam, 1987.
7 J. A. Pask and A. G. Evans (eds.), Ceramic Microstructure '86: Role of Interfaces, Plenum, New York, 1987.

8 E. A. Almond, C. A. Brookes and R. Warren (eds.), Science of Hard Materials, Inst. Phys. Conf. Ser., 75, Adam Hilger, Bristol, 1986.
9 G. S. Upadhyaya (ed.), Sintered Metal-Ceramic Composites, Elsevier, Amsterdam, 1984.
10 K. L. Mittal (ed.), Adhesion Measurement of Thin Films, Thick Films, and Bulk Coatings, ASTM Spec. Tech. Publ. 640 (1978).
11 B. V. Derjaguin, Recent Advances in Adhesion, Gordon and Breach, New York, 1971, p. 513.
12 E. D. Hondros, in R. A. Rapp ed., Techniques of Metals Research Series, Vol. IV 2, Wiley Interscience, New York, 1970, pp. 293-348.
13 J. W. Hancock, I. L. Dillamore and R. E. Smallman, 6th Planver Seminar, Planvee, Reutte, 1968, p. 467.
14 K. E. Easterling, H. F. Fischmeister and E. Navara, Powder Metall., 16 (1973) 128.
15 M. Nicholas, J. Mater. Sci., 3 (1968) 571.
16 I. G. Palmer and G. C. Smith, Proc. 2nd Bolton Landing Conf. on Oxide Dispersion Strengthening, Gordon and Breach, New York, 1968, p. 351.
17 M. McLean and E. D. Hondros, J. Mater. Sci., 6 (1971) 19.
18 J. T. Klomp, in J. A. Pask and A. G. Evans (eds.), Ceramic Microstructure '86: Role of Interfaces, Plenum, New York, 1987, p. 307.
19 J. T. Klomp, in Y. Ishida (ed.), Fundamentals of Diffusion Bonding, Elsevier, Amsterdam, 1987, p. 3.
20 H. F. Fischmeister, E. Navara and K. E. Easterling, Met. Sci. J., 6 (1972) 211.
21 J. L. Smialek and R. Browning, NASA Tech. Memo. 87168, 1987.
22 L. Pauling, General Chemistry, 3rd edn., Freeman, San Francisco, 1970, p. 118.
23 E. D. Doyle, J. G. Horne and D. Tabor, Proc. R. Soc. London Ser. A, 366 (1979) 173.
24 A. M. Stoneham, Appl. Surf. Sci., 14 (1982-1983) 249.
25 A. M. Stoneham and P. W. Tasker, Harwell Rep. T11166 (1985).
26 W. K. Kingery, H. K. Bowen and R. Uhlmann, Introduction to Ceramics, Wiley, New York, 1976, p. 210.
27 J. T. Klomp, in E. A. Giess, K. N. Tu and D. R. Uhlmann (eds.), Electronic Packaging Materials Science, Mater. Res. Soc. Symp. Proc., 40 (1985) 381.
28 J. E. McDonald and J. G. Eberhart, Trans. AIME, 233 (1965) 512.
29 S. V. Pepper, J. Appl. Phys., 47 (1976) 801.
30 K. H. Johnson and S. V. Pepper, J. Appl. Phys., 53 (1982) 6634.
31 H. F. Fischmeister, in J. A. Pask and A. G. Evans (eds.), Ceramic Microstructure '86: Role of Interfaces, Plenum, New York, 1987, p. 1.
32 A. B. Anderson, S. P. Mehandru and J. L. Smialek, J. Electrochem. Soc., 132 (1985) 1695.
33 A. B. Anderson, C. Ravimohan and S. P. Mehandru, J. Electrochem. Soc., 134 (1987) 1789.
34 S. G. Louie and M. L. Cohen, Phys. Rev. B, 13 (1976) 2461.
35 S. G. Louie, J. G. Chelikowsky and M. L. Cohen, Phys. Rev. B, 15 (1977) 2154.
36 K. Kunc and R. M. Martin, Phys. Rev. B, 24 (1981) 3445.
37 C. G. Van de Walle and R. M. Martin, J. Vac. Sci. Technol. B, 3 (1985) 1256.
38 P. Bloechl and O. K. Anderson, to be published.

39 R. Car and M. Parrinello, *Phys. Rev. Lett.*, 55 (1985) 2471.

40 A. Rahman, in J. W. Halley (ed.), *Correlation Functions and Quasiparticle Interactions in Condensed Matter*, NATO Advanced Study Ser. 35, Plenum, New York, 1977.

41 W. Kohn and L. J. Sham, *Phys. Rev.*, 140 (1965) A1133.

42 G. Kalonji, *J. Phys. Paris, Colloq. C4*, 46 (1985) 249.

43 C. Wagner, *J. Electrochem. Soc.*, 103 (1956) 571.

44 J. L. Meijering, in H. Herman (ed.), *Advances in Materials Research*, Vol. 5, Wiley Interscience, New York, 1971, p. 1.

45 A. G. Khachaturyan, *Theory of Structural Transformations in Solids*, Wiley, New York, 1983.

46 J. Matthews, *Epitaxial Growth*, Vol. 1, Academic Press, New York, 1975.

47 A. P. Sutton and R. W. Balluffi, *Acta Metall.*, 35 1987

48 H. Gleiter and H. S. Fecht, *Acta Metall.*, 33 1985 577.

49 R. H. Hoel, *Surf. Sci.*, 169 (1986) 317

50 C. A. M. Mulder and J. T. Klomp, *J. Phys. Paris, Colloq. C4*, 46 (1985) 111.

51 D. B. McWhan, *Mater. Res. Soc. Symp. Proc.*, 37 (1985) 493.

52 M. Florjancic, W. Mader, M. Ruhle and M. Turwitt, *J. Phys. Paris, Colloq. C4*, 46 1985 129

53 J. Budai, P. D. Bristowe and S. L. Sass, *Acta Metall.*, 31 1983 699.

54 J. Vitek and M. Ruhle, *Acta Metall.*, 34 (1986) 2095.

55 A. Bourret, *J. Phys. Paris, Colloq. C4*, 46 (1985) 27.

56 R. W. Balluffi, M. Ruhle and A. P. Sutton, *Mater. Sci. Eng.*, 89 (1987) 1.

57 J. M. Gibson, R. T. Tung, J. M. Phillips and R. Hull, *J. Phys. Paris, Colloq. C4*, 46 (1985) 369.

58 W. Mader and M. Ruhle, *Acta Metall.*, in press.

59 W. Mader, *Mater. Res. Soc. Symp. Proc.*, 82 (1987) 403.

60 R. C. Pond, *J. Microsc.*, 135 (1984) 213.

61 M. Backhaus-Ricoult and H. Schmalzried, *Ber. Bunsenges. Phys. Chem.*, 89 (1985) 1323.

62 W. W. Mullins and R. F. Sekerka, *J. Appl. Phys.*, 34 1963 323, 35 (1964) 444.

63 M. Backhaus-Ricoult, *Ber. Bunsenges. Phys. Chem.*, 91 1987 684.

64 K. Burger and M. Ruhle, *Adv. Ceram.*, in press.

65 K. Burger, W. Mader and M. Ruhle, *Ultramicroscopy*, 12 1987 1

66 M. Ruhle, K. Burger and W. Mader, *J. Microsc. Spectrosc. Electron.*, 11 (1986) 163.

67 M. Ruhle, M. Backhaus-Ricoult, K. Burger and W. Mader, in J. A. Pask and A. G. Evans (eds.), *Ceramic Microstructure 86: Role of Interfaces*, Plenum, New York, 1987, p. 295.

68 J. A. Wasynczuk and M. Ruhle, in J. A. Pask and A. G. Evans (eds.), *Ceramic Microstructure 86: Role of Interfaces*, Plenum, New York, 1987

69 M. Wittmer, *Mater. Res. Soc. Symp. Proc.*, 40 (1985) 393

70 B. Derby and E. R. Wallach, *J. Mater. Sci.*, 17 1982 49; 18 (1984) 427.

MATERIALS
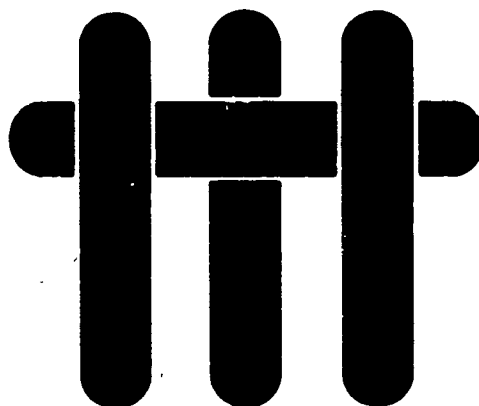
AN EXPERIMENTAL STUDY OF

THE FRACTURE RESISTANCE OF

BIMATERIAL INTERFACES

H. C. Cao and A. G. Evans

Materials Department
College of Engineering
University of California, Santa Barbara
Santa Barbara, California 93106

AN EXPERIMENTAL STUDY OF THE FRACTURE RESISTANCE OF BIMATERIAL INTERFACES

H.C. CAO and A.G. EVANS

Materials Department, College of Engineering, University of California, Santa Barbara, CA 93106, U.S.A.

Received 10 August 1988; revised version received 14 September 1988

The fracture resistance of a model bimaterial interface has been measured for a wide range of phase angles: the measure of the relative crack face sliding and opening displacement near the crack tip. These experiments have revealed that the critical strain energy release rate increases with increase in phase angle, especially when the crack opening becomes small. The results are consistent with a model based on crack surface contact associated with non-planarity of the fracture interface.

1. Introduction

When a component containing bimaterial interfaces is subjected to residual and/or external stresses, failure may occur either within one of the constituents or along the interface (e.g., Turwitt et al., 1985; Berndt and McPherson, 1981; Twentyman and Homcock, 1981). The mechanism of failure depends on the geometry of the specimen, the loading, the matrix toughness and interface fracture resistance. For example, decohesion of thin films on brittle substrates often initiates preferentially at the specimen edge and initially propagates inward along the interface. However, after extending for a length of the order of film thickness, the decohesion crack is frequently observed to deviate into the brittle substrate (Drory et al., 1988). This phenomenon and others (Evans and Hutchinson, 1989) cannot be explained by using the normal assumption that the fracture resistance of the interface is governed by a constant value of the energy release rate.

The present study consists of an experimental approach that elucidates some important features concerning the above debonding and interface fracture resistance issues. Specifically, the debonding phenomenon is examined on a model interface using composite beams and a composite cylinder. Measurements of the critical energy release rate $\mathcal{G}_c$ are made on geometries having a spectrum of

"mixed mode" loading Ψ. Following Rice (1988), the phase angle Ψ is defined through a complex stress intensity factor associated with an interface crack,

$$K = |K| e^{i\Psi} \tag{1}$$

where $|K|$ is the modulus of K, and $\Psi = \tan^{-1}[\mathrm{Im}(K\lambda)/\mathrm{Re}(K)]$. A failure locus, $\mathcal{G} - \Psi$, is thereby established and a crack surface contact zone model developed by Evans and Hutchinson (1989) is used to rationalize the experimental results. Finally, the important transition between interfacial cracking and kinking out of the interface is discussed.

2. Theoretical background

2.1 Energy release rate

Four specimen geometries are used in the present study as depicted in Fig. 1: symmetric and asymmetric cantilever beam specimens, a four-point flexure beam and a cylindrical tensile specimen. The strain energy release rate for a symmetrical, homogeneous *double cantilever beam* can be approximated by beam theory provided that the crack length/beam thickness ratio is large (Fig. 1a). Then, the strain energy stored within each
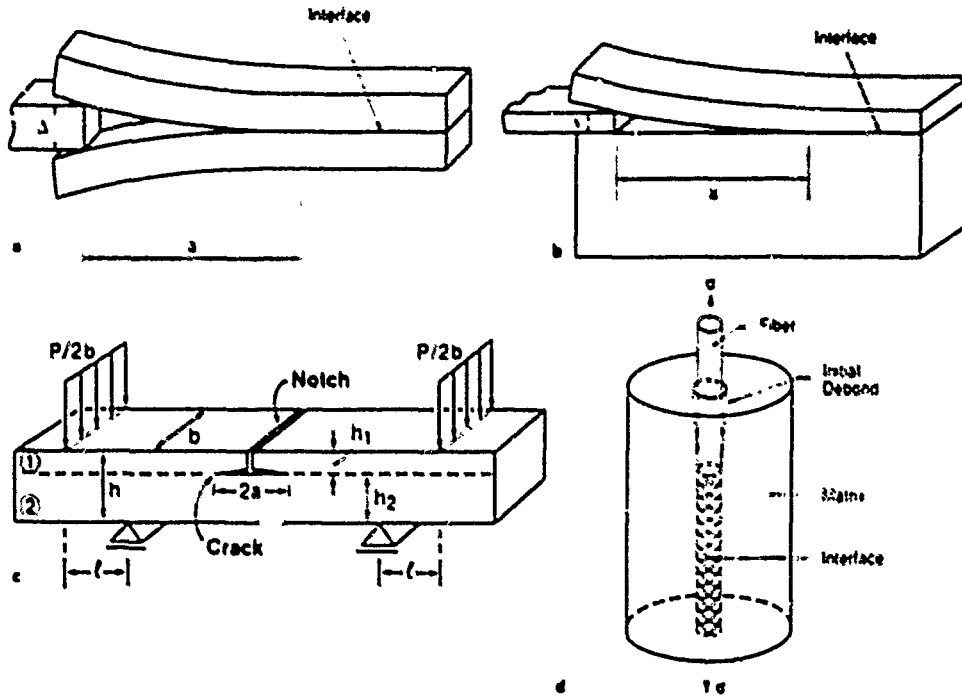
Fig. 1. Specimen geometries for interface fracture resistance measurements: (a) symmetric double cantilever beam, (b) asymmetric cantilever beam, (c) four-point flexure, and (d) composite cylinder.

beam, U, ignoring the shear contribution, is given by (Crandall et al., 1978),

$$U = \frac{2P^2a^3}{Eh^3b} \tag{2a}$$

where E is Young's modulus, h the beam thickness, a the crack length, P the applied load, and b the beam width. Therefore, the strain energy release upon crack extension becomes,

$$\mathcal{G} = \frac{2}{b}\left(\frac{\partial U}{\partial a}\right)_P = \frac{12P^2a^2}{Eb^2h^3} \tag{2b}$$

The strain energy release rate can also be expressed in terms of the displacement at the point of loading, Δ, (Tada et al., 1973) as,

$$\mathcal{G} = \frac{3E\Delta^2h^3}{16a^4} \tag{3}$$

where

$$\Delta = \frac{4Pa^3}{\sqrt{3}\,Eh^3b}.$$

In the asymmetric *beam* specimen (Fig. 1b), only the strain energy in the peeled arm contributes to $\mathcal{G}$. Approximating the arm by a built-in beam with a vertical force P exerted at the end, the strain energy release rate is given by (Tada et al., 1973),

$$\mathcal{G} = \frac{6P^2a^2}{E_1b^2h^3} = \frac{3E_1\Delta^2h^3}{8a^4} \tag{4}$$

where E_1 refers to the cantilever beam.

In the case of the *four-point flexure* specimen, a constant moment exists within the inner loading points (Fig. 1c). Consequently, the energy release

rate of the interface crack exhibits steady-state behavior, wherein $\mathcal{G}$ is independent of the crack length (Charalambides et al., 1989a). Again, by using beam theory, the strain energy release rate can be evaluated by subtracting the strain energy stored in a unit section of composite beam from the strain energy stored in the uncracked beam,

$$\mathcal{G} = \frac{M^2}{2E_1 b}\left(\frac{E_1}{E_2 I_2} - \frac{1}{I_c}\right)$$ (5)

where the subscript c refers to the composite beam. $M = Pl/2$, with l being the spacing between the inner and outer loading lines: I_2 and I_c refer to the moment of inertia of the uncracked and composite beams respectively.

$$I_2 = bh_2^3/12$$

$$I_c = bh_1^3/12 + \lambda bh_2^3/12$$

$$+ \lambda bh_1 h_2(h_1 + h_2)^2/4(h_1 + \lambda h_2)$$

with $\lambda = E_2/E_1$.

For the axisymmetric tensile test, the energy release rate again reaches a steady state, when the debond length is greater than the fiber diameter (Fig. 1d). The steady-state energy release rate can be evaluated by subtracting the strain energy stored in a section of the "fiber" from that in the lower portion of the composite, leading to (Charalambides and Evans, 1989),

$$\mathcal{G} = \frac{t^2\lambda R}{4E_f[\lambda + f/(1-f)]}$$ (6)

where R is the "fiber" radius, t the applied stress, f the "fiber" volume fraction and E_f the "fiber" modulus.

2.2. The phase angle of loading

For each of the specimen geometries described above, the phase angle of loading Ψ develops an unique value, independent of crack length, when the crack is long compared with other characteristic dimensions (for example, the beam thickness or the fiber radius). The steady-state Ψ has been calculated for each geometry using a finite ele-
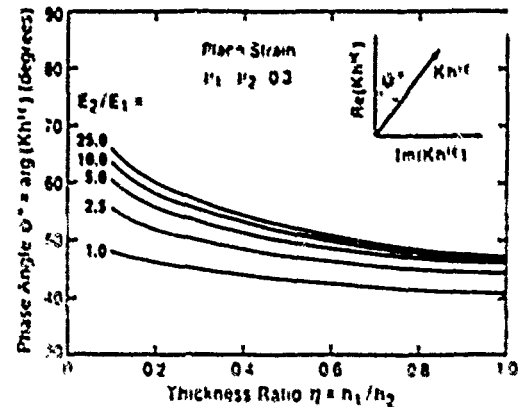

Fig. 2 The steady-state non-dimensional phase angle Ψ^* for an interface crack in a flexure specimen, with respect to thickness ratio h_1/h_2 and modulus ratio E_2/E_1 for fixed Poisson's ratio, $\nu_1 = \nu_2 = 0.3$.

ment procedure described elsewhere (Matos et al., 1989). The salient results are described below.

An elastically homogeneous symmetric DCB specimen is obviously subject to mode I($\Psi = 0$). However, when the upper and lower beams are either of different thickness or made from different materials, the phase angle deviates from zero. The result most relevant to the present study gives a phase angle of 36° when the beam thickness ratio is 0.1 and the upper and lower beams have similar modulus and Poisson's ratio.

The phase angle for the flexure specimen varies with the thickness ratio, h_1/h_2, and modulus ratio, λ, as depicted in Fig. 2. The phase angles relevant to the present specimens ($E_2 = E_1$) are $\Psi = 40°$ when $h_1/h_2 = 1$, and $\Psi = 49°$ when $h_1/h_2 = 0.1$.

For the axisymmetric specimen, the phase angle varies with the "fiber" volume fraction and the elastic constants of the system (Fig. 3). For present purposes, it is noted that, when the "fiber" volume fraction is small and when "fiber" and matrix have similar modulus and Poisson's ratio, the phase angle is $-70°$.

2.3 The interfacial zone

When the crack lies on one interface of a thin layer, thickness h, bounded to two thick bodies (Fig. 4), the strain energy release rate can be
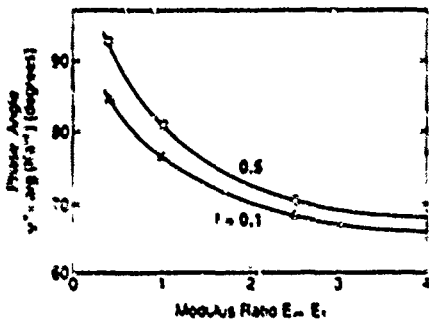
Fig. 3. The steady-state non-dimensional phase angle for a composite cylinder subject to tension.

estimated by ignoring the layer if its thickness, h, is small compared with other specimen dimensions and crack length. However, the phase angle for the near tip stress field is modified from that of the asymptotic field. It has been shown by Suo and Hutchinson (1989) that the modified phase angle can be related to the remote quantity through

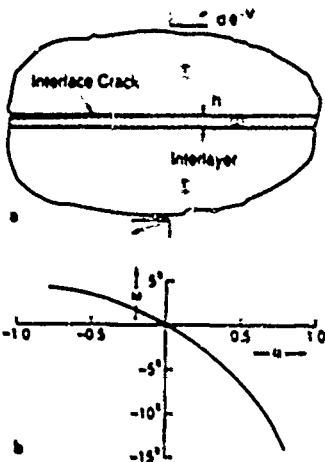$$\Psi^{tip} = \Psi - \epsilon \ln h + \omega(\alpha, \beta) \qquad (7)$$

Fig. 4. (a) A three-layer composite with a semi-infinite crack on an interface. (b) Rotation parameter ω caused by the presence of a thin layer for $\beta = 0$. (However, ω is insensitive to β.)

where (Dundurs, 1969),

$$\alpha = \frac{\mu_1(1-\nu_2) - \mu_2(1-\nu_1)}{\mu_1(1-\nu_2) + \mu_2(1-\nu_1)}$$

$$\beta = \frac{1}{2} \frac{\mu_1(1-2\nu_2) - \mu_2(1-2\nu_1)}{\mu_1(1-\nu_2) + \mu_2(1-\nu_1)} \qquad (8)$$

with μ being the shear modulus, ν Poisson's ratio and the subscripts 1 and 2 now referring to the thin layer and the material on the other side of the cracked interface, respectively, and

$$\epsilon = \frac{1}{2\pi} \ln\left(\frac{1-\beta}{1+\beta}\right)$$

It has been pointed out by Suo and Hutchinson that ϵ and ω are insensitive to β. For the very thin layers used in this study ($h \approx 25\,\mu m$), and using micrometers as the unit for h, Ψ^{tip} and Ψ differ by $\sim 5°$ (Fig. 4). The interface failure locus $\mathcal{G}_c(\Psi^{tip})$, can thus be recovered from the remote quantities $\mathcal{G}_c(\Psi)$, provided that the layer thickness is small compared with other specimen dimensions.

3. Experimental

3.1 Material

One objective of the present experimental study is to observe the sequence of events associated with bimaterial interface failure under general mixed mode loading. A translucent material facilitates such observations. For this purpose, glass plates and rods have been selected with surfaces that are smooth and uniform. An aluminum alloy is chosen as another component because it has similar elastic properties to glass, thus eliminating complexities arising from elastic mismatch. However, the thermal expansion mismatch between the two is significant, allowing residual stress effects to be explored. The bonding agent used to join the specimens is a thermoplastic adhesive which softens at 74°C and flows at 135°C. At room temperature, the material is translucent and very brittle. In addition, it can be easily dissolved into acetone.
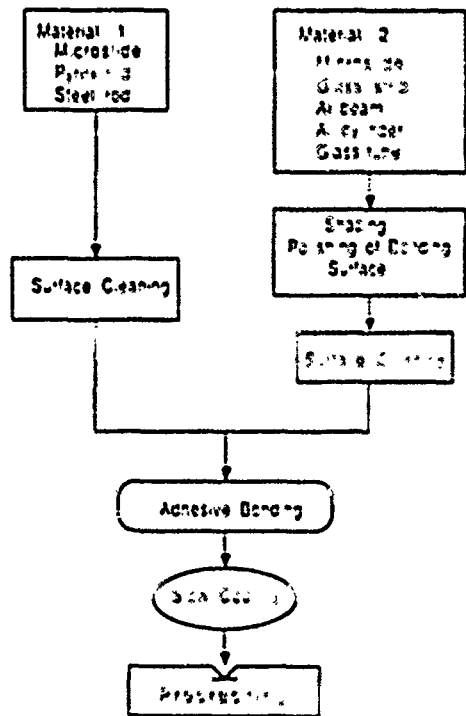
Fig. 6. Experimental setup for precracking the DCB specimen

Fig. 5. Flowchart illustrating specimen preparation procedures.

3.3 Testing

For the *cantilever beam* specimens, initial cracking is achieved by inserting a wedge between the beams, clamped at the other end (Fig. 6). The clamp exerts a compressive pressure on the interface and arrests the crack. The clamp is then gradually removed and the critical crack length for interface crack extension measured upon re-inserting the wedge.

The *flexure* specimen is precracked using three-point bending (Fig. 7). For this purpose, a scratch is inscribed onto the glass before bonding. Then, upon three-point loading, the scratch propagates toward the interface in an unstable fashion and branches into the interface. However, because the driving force for the interface crack diminishes

3.2 Specimen preparation

The bonding procedure (Fig. 5) has been selected to maximize reproducibility. Furthermore, to assure crack surface uniformity, the bonding surfaces of aluminum are first polished with $\frac{1}{4}$ μm diamond paste resulting in a roughness of ~0.1 μm. Each of the constituents is then immersed into acetone for 30 min and subsequently cleaned in an ultrasonic cleaner. Bonding is conducted on a hot plate at controlled temperature. Bubbles within the adhesive are driven out by shearing and establishing bonding from one edge. These procedures are followed for both the planar and cylindrical geometries. The cooling rate following bonding is also controlled so that the residual stresses remain constant among specimens. The amplitude of the residual stress is assessed separately (Appendix).
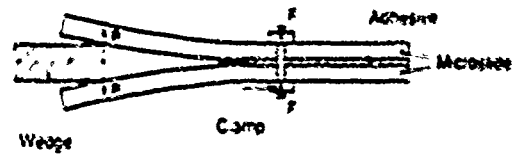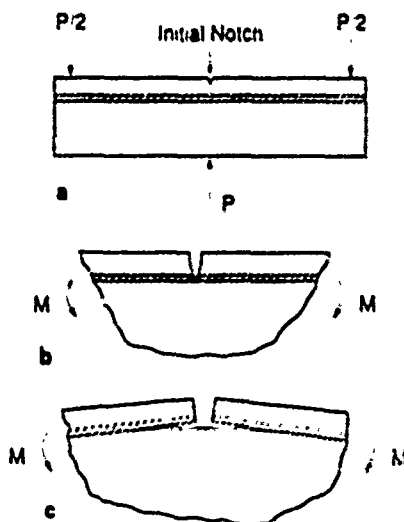

Fig. 7 Procedures for creating decohesion pre-crack in flexure specimen using three-point bending: (a) three-point bend of initial notched beam, (b) crack terminating at the interface, and (c) decohesion at adhesive/substrate interface.
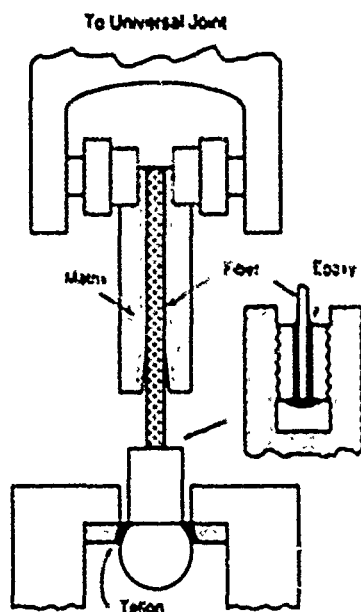
To Universal Joint

Glass   Foil   Epoxy

Teflon

Fig. 8. Schematic of tensile fixture for composite cylinder testing.

smooth transition of the bond coupled with the viscous deformation of the epoxy alleviate the stress concentration at the grips. The rod is connected to a hardened steel ball within a spherical washer having a thin layer of Teflon between the contacts. This arrangement provides a moment free connection between the two ends.

4. Observations and measurements

4.1 The failure locus

For the cantilever beam tests, the difference in the reflection of light by the bonded interface and the decohered region makes the crack tip readily visible. Such optical monitoring of debonding reveals that, when the crack becomes critical, it jumps, despite the decreasing nature of the energy release rate as crack extends. The critical energy release rate for crack initiation, $\mathcal{G}_i$, and crack arrest, $\mathcal{G}_a$, thus differ. To validate the test procedure, the crack initiation energy release rate $\mathcal{G}_i$ was measured using different wedges (Fig. 9). The results exhibit the required insensitivity to the wedge opening within experimental uncertainty ($\mathcal{G}_i = 7 \pm 2$ Jm^{-2}). For the flexure and tensile tests, crack length measurement is not important because the energy release rate and phase angle are constant. However, an LVDT is used to monitor the displacement, because debond growth is then clearly delineated by a displacement discon-
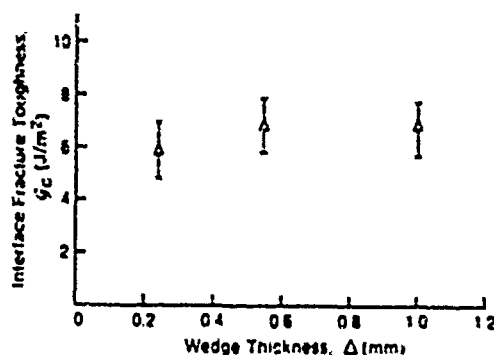
as the crack extends away from the central plane, the crack arrests. The crack arrest length is controlled to be approximately 10 times the thickness of the cracked beam. The precracked beam is then transferred to the four-point loading fixture, with the arrested crack tip located between the inner loading points, whereupon steady-state conditions exist. For this geometry, frictional forces at the loading points have a profound effect on the values of $\mathcal{G}$ inferred from the critical loads (Charalambides et al., 1989a). Consequently, friction is first minimized by inserting a thin layer of Teflon tape between the specimen and the loading points. The amplitude of the resultant friction is then measured from the hysteresis in the elastic compliance.

For the tensile specimen consisting of a glass rod bonded into a glass tube, an initial debond is created by partially dissolving the adhesive in acetone, using ultrasonic vibration to cause the solvent to penetrate uniformly. The specimen is then loaded in a tensile fixture (Fig. 8), with the glass rod bonded to a set screw using epoxy. The

Fig. 9. Interface fracture resistances obtaining using asymmetric cantilever specimens ($\Psi = 36°$) for three wedge thicknesses.
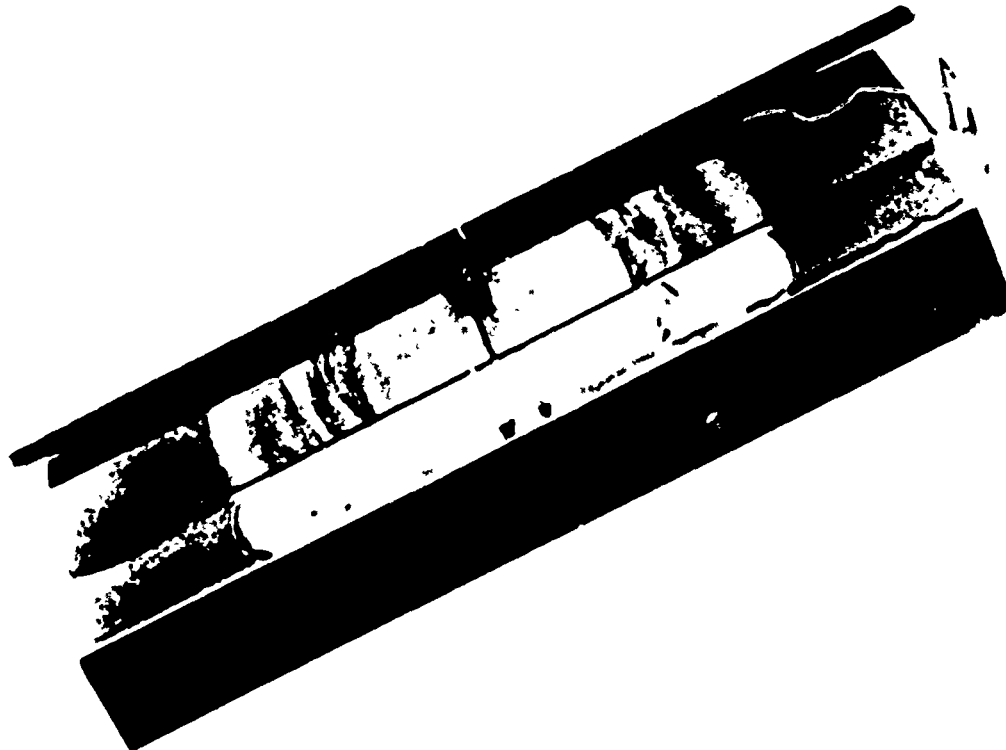
Fig. 10. An optical view of the interface of a flexure specimen after precracking and subsequent extension of interface crack.

tinuity. A typical specimen after a flexure test (Fig. 10) provides clear evidence of markings associated with initial growth and crack arrest.

The measurements of $\mathcal{G}_c$ obtained from each of the above tests are combined in Fig. 11 to establish the failure locus for the interface by applying the values of the phase angle described in Section 2. For the Al/glass specimens, the residual stress was taken into account, as described in the Appendix. It is apparent that $\mathcal{G}_c$ is larger for the glass/polymer interface than for the Al/polymer interface and the $\mathcal{G}_c$ increases as Ψ increases, becoming quite large when Ψ approaches $\pi/2$.

4.2 The crack extension path

Another important issue concerns the crack extension path in a three-layer composite (Fig.

12). When the crack is subject to opening mode loading ($\Psi = 0$), crack extension is observed to occur within the interlayer. However, when subject to mixed mode loading, the crack always propagates within the interface. Furthermore, the crack selects the upper interface when $\Psi < 0$ and the lower interface when $\Psi > 0$. Specifically, for asymmetric cantilever specimens in which $\Psi < 0$, the crack extends on the upper interface, whereas in the flexure specimen, for which $\Psi > 0$, the crack path is along the lower interface. Additionally, and most insightful, when a flexure test is interrupted and a cantilever force applied, the interface crack path abruptly changes (Fig. 12c). These results provide definitive evidence that the crack trajectory and thus the preferred fracture interface is largely dominated by the sign of the phase
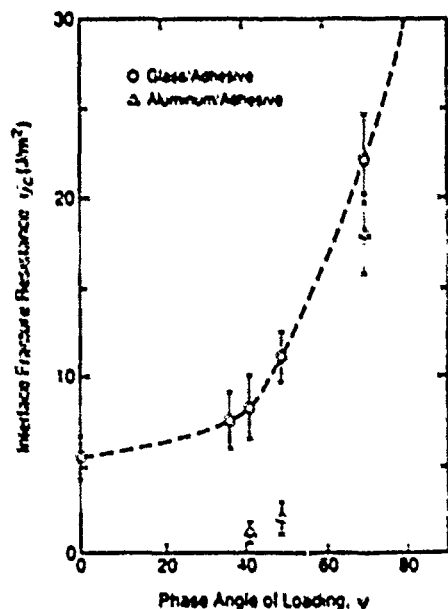
Fig. 11. Interface fracture resistance with respect to phase angle.

where H is the amplitude of the interface roughness, L the wavelength, $\mathcal{G}_0$ is the fracture resistance of the interface at $\Psi = 0$ and E is Young's modulus. Trends in crack shielding $\Delta\mathcal{G}/\mathcal{G}_0$ with Ψ predicted by the model are plotted on Fig. 13: the present experimental results are juxtaposed. Comparison indicates that agreement between theory and experiment implies a magnitude of χ of order 10, reasonably consistent with the characteristics of the present systems ($E = 2 - 70$ GPa, $H = 0.1$ μm, $L = 1$ μm, $\mathcal{G}_0 = 6$ Jm^{-2}).

angle, at least when the interlayer material has a relatively low toughness.

5. Comparison with theory

The only well developed model that describes the trends in $\mathcal{G}_c$ with Ψ invokes a non-planar interface and examines the effect of the contact provided by non-planarity on the shielding of the tip from the applied loads (Evans and Hutchinson, 1989). This model has several simplifying features, such as an idealized interface morphology (Fig. 13a), homogeneous elastic properties, and freely sliding contacts (no friction). Nevertheless, the model identifies the existence of a non-dimensional material parameter χ that profoundly influences the trend in $\mathcal{G}_c$ with Ψ (Evans and Hutchinson, 1988),

$$\chi = EH^2/\mathcal{G}_0 L$$
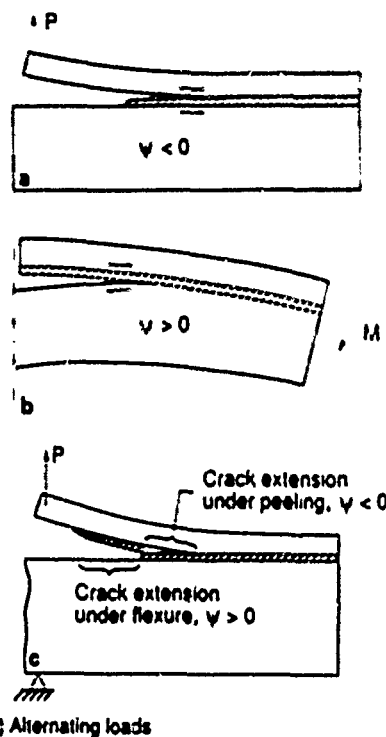

c) Alternating loads

Fig. 12. (a) Crack extension path taken by the interface crack in peel specimen. (b) Crack path in flexure specimen. (c) Schematic illustrating crack extension path change when a flexure specimen is subsequently loaded at end of beam, noting interface crack kinking into adhesive and extending on the other interface of a thin interlayer structure, when the phase angle changes sign.
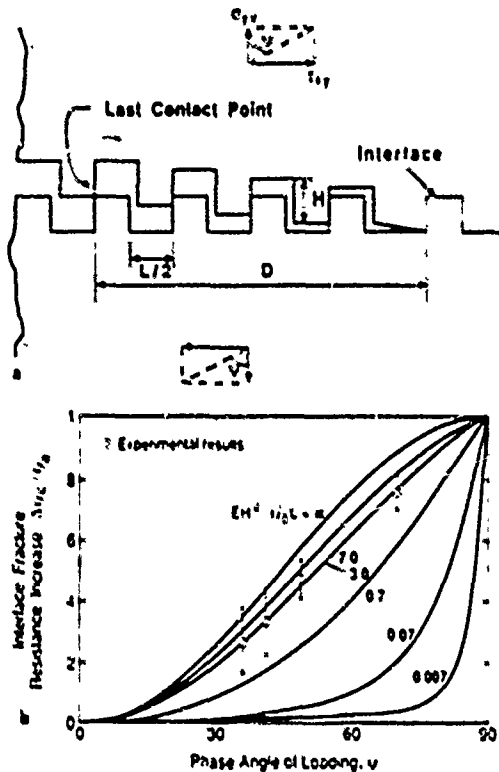
Fig. 13. (a) Schematic of contact zone illustrating crack face contact upon transverse displacement. (b) Trends in crack shielding $\Delta \mathcal{G}_c / \mathcal{G}_0$ with phase angle of loading Ψ for various values of the contact zone parameter χ. The experimental results are also shown.

6. Conclusions

Mixed mode fracture conditions have been established for a model bimaterial interface by introducing interface cracks into different specimen geometries. The interface fracture resistance has been found to increase with increase in phase angle of loading. A simple contact zone model predicts trends consistent with the experimental results. However, the trend could also be influenced by other phase angle dependent phenomena, such as blunting and plasticity.

The fracture trajectory has been shown to depend on the sign of the phase angle, and can be moved from one interface to another by changing the sign of Ψ while $\mathcal{G}$ remains constant. Furthermore, the crack remains in the bond layer when $\Psi = 0$. Measurements of the interface fracture resistance must, therefore, be cognizant of the sign of Ψ and of the fracture interface involved.

Appendix

Residual stress effects in beams

Thermal residual stresses arise within composite beams when cooled from the processing temperature. The differential shrinkage between the two beams deforms the composite beam into a curved shape. The curvature of the beam can be measured using a profiler and the residual stress computed following a beam theory solution. Specifically, when a profiler travels a distance, d, on the beam surface, a maximum displacement, δ, occurs at the center point such that the misfit stress is

$$\sigma_t = \frac{4 E I_c \delta}{b h_1 h_2 d^2} \tag{A.1}$$

where h_1 and h_2 are the thickness of the upper and lower beams respectively and I_c is the moment of inertia for the composite beam.

When an external load is applied (Fig. 1c), both the applied and residual stress contribute to the energy release rate upon crack propagation, such that (Charalambides et al., 1989b)

$$\mathcal{G} = \frac{1 - v_1}{2 E_1 b} \left[\sigma_t^2 h_1^2 \left(\frac{1}{A_1} - \frac{1}{A_c} \right) + \frac{M^2}{\lambda I_c} \right.$$
$$\left. - \frac{(M - \sigma_t h_1 h_2)^2}{I_c} \right] \tag{A.2}$$

where $A_1 = b h_1$, $A_c = (h_1 + \lambda h_2) b$, and $M = b h_1^2 \sigma_m / 6$ is the applied moment with σ_m being the maximum bending stress over the uncracked beam. Rearranging gives:

$$\mathcal{G} = \frac{1 - v_1}{2 E_1} \sigma_m^2 h_1 \Phi(\xi, \eta, \lambda) \tag{A.3}$$

where

$$\phi(\xi, \eta, \lambda) = \frac{\xi}{3} + \eta^2\left(1 - \frac{1}{1+\lambda\xi}\right)$$

$$+ \frac{4\xi^3\eta - \xi^2/3 - \xi^2\eta^2}{1 + \lambda\xi^3 + 3\lambda\xi(1+\xi)^2/(1+\lambda\xi)}$$

with $\xi = \sigma_t/\sigma_m$ and $\eta = h_1/h_2$. Consequently, by evaluating σ_t from the bending and determining σ_m from the applied load at crack extension, $\mathcal{G}$ can be determined from eq. (A.3). This procedure is used for the Al/glass specimens.

The phase angle is also modified when residual stress is present (Charalambides et al., 1989b). However, for the relative stress levels arising during the present experiments, ($\xi = 0.2$), this effect is small.

References

Berndt, C.C. and R. McPherson (1981), in: J.A. Pask and A.G. Evans, eds., *Surface and Interfaces in Ceramic and Ceramic / Metal Systems*, Plenum, New York, p. 619.

Charalambides, P.G. and A.G. Evans (1989), Interface fracture in brittle matrix composites, *J. Amer. Ceram. Soc.*, May.

Charalambides, P.G., J. Lund, A.G. Evans and R.M. McMeeking (1989a), A test specimen for interface fracture resistance measurement, *J. Appl. Mech.*, March.

Charalambides, P.G., H.C. Cao, J. Lund and A.G. Evans (1989b), *Acta Metall.*, to be published.

Crandall, S.H., N.C. Dahl and T.J. Lardner (1978), *An Introduction to the Mechanics of Solids*, McGraw-Hill, New York, 2nd edn.

Drory, M.D., M.D. Thouless and A.G. Evans (1988), On the decohesion of residually stressed thin films, *Acta Metall.* 36, 2019.

Dundurs, J. (1969), *Mathematical Theory of Dislocations*, ASME, New York, p. 70.

Evans, A.G. and J.W. Hutchinson (1989), Effects of non-planarity of the mixed mode fracture resistance of bimaterial interfaces, *Acta Metall.* 37, 909.

Marshall, D.B. (1984), An indentation method for measuring matrix-fiber frictional stresses in ceramic composites, *J. Amer. Ceram. Soc.* 67(12) C259

Matos, P., P.G. Charalambides, M.D. Drory and R.M. McMeeking (1989), *Int. J. Fract.*, in press.

Rice, J.R. (1988) *J. Appl. Mech.* 55, 98.

Suo, Z. and J.W. Hutchinson (1989), *Mater. Sci. Eng.*, in press.

Tada, H., P.C. Paris and G.R. Irwin (1973), *The Stress Analysis of Cracks Handbook*, Del Research Corporation.

Turwitt, M., G. Elssner and G. Petzow (1985), Manufacturing and mechanical properties of interfaces between sapphire and Niobium, *J. Phys. C4*, 123.

Twentymen, M.E. and P. Homcock (1981), in: J.A. Pask and A.G. Evans, *Surfaces and Interfaces in Ceramic and Ceramic/Metal Systems*, Plenum, NY, p. 513